UNIVERSITÀ DEGLI STUDI FIRENZE

DOTTORATO DI RICERCA

INTERNATIONAL DOCTORATE IN STRUCTURAL BIOLOGY

CICLO XXVII

COORDINATORE Prof. Claudio Luchinat

Development of new NMR methods to study intrinsically disordered proteins

Settore Scientifico Disciplinare CHIM/03

Dottorando
Dott. Tomáš Hošek

Tutore
Prof. Isabella C. Felli

CoTutore
Prof. Roberta Pierattelli

Coordinatore
Prof. Claudio Luchinat

Anni 2012/2014

This thesis has been approved by the University of Florence, the University of Frankfurt and the Utrecht University

Universiteit Utrecht

CERM Firenze

Frankfurt am Main

Acknowledgements

This thesis would not have been possible without the support of many people.
I am deeply grateful for the great opportunities that have been given to me by the IDPbyNMR project in the frame of the Marie Skodowska-Curie Actions (MSCA). Joining this program introduced me to the vast international scientific community and more importantly it allowed me to personally interact with the top scientists in diverse research areas. The meetings organized within the IDPbyNMR network broaden my knowledge in many directions, ranging from bioinformatics and different biophysical techniques to personal presentation skills. The IDPbyNMR doesnt let me grow only as a scientist but more importantly also as a person.
I would like to express my sincere gratitude to my supervisor prof. Isabella C. Felli and co-supervisor prof. Roberta Pierattelli for constant support, guidance, teaching and lastly for their great patience with my "peculiar features". I would like to thank to Dr. Bernhard Brutscher who hosted me and took care of me for two month during my secondments in CNRS-IBS.
I would like also to thank many people who I was happy to meet during my doctorate years in CERM for their help, discussions and hilarious moments: Eduardo, Alessandro, Angelo, Daniela, Magda, Serena, Karolina, Maciek, Marcela, all colleagues who became my dear friends, and many others. During my doctoral studies I also had opportunity to guide a master student, Michele, who actually taught me more than I taught him, so I would like to thank him for being the first person joining my "teaching trial studies".
I would like to specially thank Karel Kubíček who introduced me to the world of NMR and always offers me invaluable assistance.
I wish to thank my family for their endless support and encouragement during my entire life.

Abstract & Keywords

Abstract

The highly flexible nature of intrinsically disordered proteins (IDPs) confers them functional advantages complementary to those resulting from the presence of a stable three dimensional structure. This is becoming evident thanks to recent active research in the field of IDPs, introducing a new dimension to the structure-function paradigm to account for the existence of protein disorder. NMR thus becomes a strategic tool of investigation. Nevertheless, the particular properties of IDPs, such as high flexibility and solvent exposure of the backbone, should be taken into account when optimizing and designing NMR experiments.

During my doctorate I focused on the development of NMR methods for the characterization of IDPs and on their application to study challenging systems. The new experiments developed address the issues of the NMR studies conducted at close to physiological conditions (temperature and pH). Longitudinal relaxation enhancement techniques are exploited to increase sensitivity of exclusively heteronuclear experiments, and to study local compactness and solvent accessibility of different regions in IDPs. New high dimensional experiments for sequence specific assignment of IDPs are proposed completing the suite of 4D/5D experiments that exploit correlations involving carbonyls and nitrogens ("CON-CON" strategy).

NMR experiments, tested on α-synuclein, a well characterized IDP, were also applied for the investigation of two IDPs of viral origin that play crucial roles in carcinogenesis, the E7 protein from human Papillomavirus and the E1A protein from human Adenovirus. IDPs are extensively exploited by viruses to imitate interaction sites of the host regulatory and signaling proteins, and thus force the infected cell to create friendly environment for viral replication. However, the molecular basis of how this is achieved is still not fully understood. The atomic level characterization of the structural and dynamic properties of IDPs involved in viral infection is expected to contribute to the understanding of underlying

molecular mechanisms.

Keywords

NMR, Intrinsically disordered proteins, IDPs, ^{13}C detection, longitudinal relaxation enhancement, LRE, SLiMs, human Papillomavirus, human Adenovirus

List of Abbreviations

IDP	Intrinsically Disordered Protein
IDPR	Intrinsically Disordered Protein Regions
PTM	Post-Translational Modification
MoRFs	Molecular Recognition Features
SLiMs	Short Linear Motifs
LRE	Longitudinal Relaxation Enhancement
SNR	Signal-to-Noise Ratio
NUS	Non-Uniform Samplling
FID	Free Induction Decay
FFT	Fast Fourier Transformation
SMFT	Sparse Multidimensional Fourier Transformation
HPV	Human Papillomavirus
HAdV	Himan Adenovirus
CR	Conserved Region
pRB	Retinoblastoma protein

Contents

Acknowledgements i

Abstract & Keywords ii

List of Abbreviation iv

1 Introduction 1

1.1 Intrinsically disordered proteins (IDPs) coming on the scene 2

1.2 Functional basis of IDPs 3

1.3 IDPs and diseases 7

1.4 High resolution structural and dynamic investigation of IDPs 8

2 NMR challenges to study IDPs 9

2.1 Spectral resolution 11

2.2 Experimental sensitivity 13

2.3 Experimental time requirements 14

2.4 Sample optimization 14

2.5 High content of proline residues 15

2.6 Fingerprint spectra of an IDP 15

3 Methodological aspects 18

3.1 ^{13}C detected experiments 19

3.1.1 Homonuclear ^{13}C decoupling 20

3.2 Improving sensitivity of ^{13}C detected experiments 22

3.2.1 Starting polarization source 23

3.2.2 Longitudinal relaxation enhancement 24

3.3 Multidimensional experiments and non-uniform sampling 26

3.3.1 Analysis and inspection of multidimensional spectra 28

4 Viral IDPs **30**

4.1 Human Papillomavirus E7 protein 31

4.2 Human Adenovirus E1A protein 34

5 Results **39**

5.1 NMR methods to study IDPs 43

5.1.1 NMR spectroscopic studies of intrinsically disordered proteins at near-physiological conditions 43

5.1.2 Longitudinal relaxation properties of $^{1}H^{N}$ and $^{1}H^{\alpha}$ determined by ^{13}C direct-detected NMR experiments to study intrinsically disordered proteins (IDPs) 59

5.1.3 "CON-CON" assignment strategy for highly flexible intrinsically disordered proteins 89

5.1.4 NMR methods for the study of IDP structure, dynamics, and interactions: general overview and practical guidelines 107

5.2 Application of NMR methods to viral IDPs 209

5.2.1 The heterogeneous structural behavior of E7 from HPV 16 revealed by NMR spectroscopy 209

5.2.2 Biophysical characterization of the E7 conserved region 3 from HPV 16 223

5.2.3 The highly flexible and heterogeneous nature of HAdV E1A characterized at atomic resolution through NMR 237

6 Conclusions and Perspectives **271**

Bibliography **274**

Chapter 1

Introduction

Our current understanding of protein function is largely influenced by the vast amount of three dimensional structures of proteins deposited in the PDB.[1] The information available in the PDB contributed to the formulation of the so-called "structure function paradigm" which essentially states that the function of a protein is uniquely determined by its three dimensional structure (*Amino acid sequence* $\rightarrow$ *3D structure* $\rightarrow$ *Function*). Indeed the three dimensional structures available in the PDB contribute to explaining the molecular basis related to protein function in a variety of different cases. Moreover the information on protein structure has contributed to the design of new drugs. However, recent evidence accumulating at an increasing speed shows that proteins characterized by a stable three dimensional structure constitute only a fraction of the functional proteins, with many of them being intrinsically disordered (IDPs) or harboring intrinsically disordered regions (IDPRs). The high flexibility and the many conformations accessible to these proteins at room temperature can provide them functional advantages that are highly complementary to those deriving from the presence of uniquely defined local environments associated to a stable three dimensional structure. This idea now seems so simple and intuitive that one may ask why it took such a long time to be accepted by the scientific community. Indeed, the propensity of a protein to adopt a more compact or disordered conformation is written directly into its primary sequence as it has recently been proved through bioinformatics studies on a large number of proteins sequences. [1, 2] The initial key studies confirming the importance of protein intrinsic disorder, the variety of functional advantages deriving from protein intrinsic disorder, and the role of intrinsic disorder in several diseases are discussed in this first chapter.

[1] http://www.rcsb.org/pdb/statistics/holdings.do

1.1 Intrinsically disordered proteins (IDPs) coming on the scene

Despite the fact that proteins or protein regions without a stable 3D structure but still biologically active were repeatedly observed and reported in the literature, they were taken more as rare exceptions than as a general phenomenon. Just to name a few, nonhistone chromosomal proteins [3–5], microtubule associated protein 2 (MAP2) [6], calmodesmon [7], cAMP-dependent protein kinase inhibitor [8], tau protein [9], heat shock transcription factor [10], and α_S-casein [11], they all share the common feature of disorder being crucial for their function. A complete overview of such proteins is summarized in several recent reviews. [12, 13]

Even though several scientists pointed out that many critical cellular events cannot depend upon the precise three dimensional structural complementarity of the interacting proteins [14, 15], the importance of structural disorder has not been recognized for a long time. The situation changed at the turn of the century. The thorough investigation of the available literature data and a comprehensive bioinformatics analysis disclosed the common features of disorder in proteins. [12, 16] Paradoxically, the data available in the PDB were also very useful for this purpose as most of the bioinformatic tools to predict intrinsic disorder come from studying regions of missing electron density maps in the deposited 3D protein structures. The combination of the low content of hydrophobic residues (leading to low hydrophobic driving force for protein compaction) and high content of charged residues (leading to strong electrostatic repulsion) represent important prerequisites for the absence of a compact structure in a protein under the physiological conditions. On this basis the first neural network predictors of structural disorder were developed and applied to the Swiss Protein Database. [1, 2] Disordered regions of at least 40 consecutive amino acids were identified in more than 18% of the deposited protein sequences. [2] These findings were further supported when bioinformatic predictors of disorder, such as PONDR, were applied on the set of protein sequences from 34 complete genomes indicating that eukaryotes have a higher percentage of proteins with long disordered regions (27-41%) than archaea (4-24%) or bacteria (2-21%). [17] This ultimately leads to the formulation of the important hypothesis that naturally flexible proteins, instead of being just rare exceptions, represent a very broad class of proteins. [12, 18, 19]

Since the information about such proteins was very sparse, a multitude of different terms was used in the literature to describe these non-traditional proteins. [20] Recently it was proposed to use the term intrinsically disordered protein (IDP) or intrinsically dis-

ordered protein region (IDPR) to indicate polypeptide chains that lack a stable, unique, well defined 3D structure. Their native state resembles the denatured state of globular proteins, which enables them to carry out important biological functions, highly complementary to those generally associated to the presence of a stable 3D structure, such as signaling and regulation. As the final step of the general acceptance of the importance of IDPs, the DisProt database [21] was created to collect all scattered experimental data on IDPs. The latest release (May 2014) of DisProt contains 1539 disordered regions in 694 IDPs.[2] More recently the pE-DB [22] was also created to host, in addition to experimental data containing structural and dynamic information such as NMR and SAXS data, also the three dimensional structural ensembles that can be generated with a variety of different approaches in agreement with experimental data. [23] As the field of IDPs is in fast development this database is expected to stimulate further progress in this area.

IDPs can be described as dynamic ensembles of multiple conformations separated by low free-energy barrier, which constantly undergo non-cooperative conformational changes. Structural disorder is present in the proteins to the various extent, ranging from fully disordered proteins to partially folded pre-molten globules with various amounts of secondary/tertiary residual structure to rather compact states similar to molten globules (Figure 1.1). [24, 25]

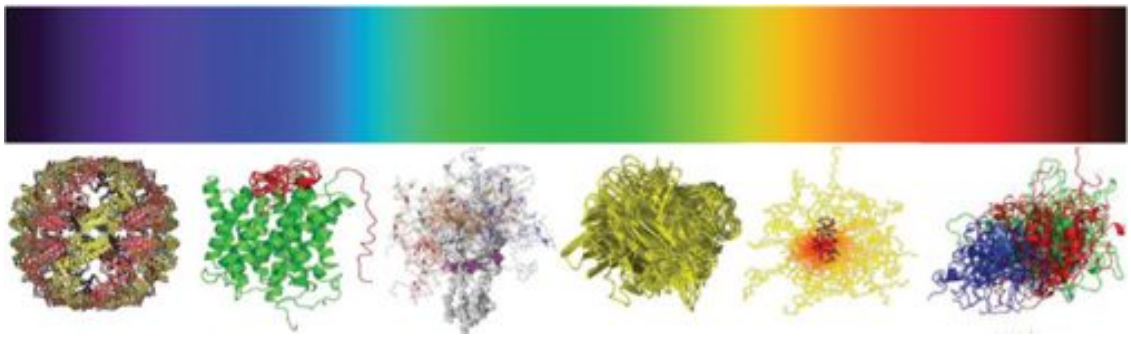


Figure 1.1. A continuous emission spectrum representing the fact that functional proteins can be disordered to different extents ranging from fully ordered to completely disordered, highly flexible proteins, with many intermediate states. Adapted from [26]

1.2 Functional basis of IDPs

Certain biological functions, such as enzyme catalysis, immunological recognition, or molecular discrimination by receptors, demand precise control of 3D structure. In con-

[2] http://www.disprot.org

trast, the conformational plasticity of IDPs and IDPRs provides them with a wide range of exceptional functional advantages over the functional modes of ordered proteins and domains, as extensively discussed in several recent reviews. [26–28] Some examples are schematically summarized hereafter:

- the ability of IDPs and IDPRs to adapt their conformations to the template provided by binding partners (*binding plasticity*);
- the possibility of overlapping binding sites due to their extended conformations;
- the ability of a single disordered region to bind several structurally diverse partners (*one-to-many binding*);
- the possibility of many different IDPRs to bind to a single structured protein (*many-to-one binding*);
- increased interaction surface per residue respect to globular proteins;
- accessible sites for post-translational modifications (PTM), a feature that allows efficient structural and functional regulation (phosphorylation, acetylation, ubiquitination, *etc.*);
- the combination of high specificity and low affinity, especially useful in regulatory interactions where turning a signal on is of the same importance as turning it off;
- production of functionally diverse forms of IDPs by alternative splicing (given the existence of multiple functions in a single disordered protein, and given that each functional element is typically relatively short, alternative splicing could readily generate a set of protein isoforms with a highly diverse set of regulatory elements). [29]

IDPs are thus related to the many functions that are difficult to be encoded by stable 3D structures such as cellular control, signaling, regulation and recognition. [24, 30–32] The various protein functions, in which intrinsic disorder is involved, are summarized in Figure 1.2. The ability of an IDP to fulfill more unrelated, or at least not obviously related, functions, sometimes described as “moonlighting ability”, represents a possible way to increase the organisms complexity without expensive expansion of the genome size. [33, 34]

The structural flexibility of IDPs and IDPRs allows them to participate in various, at first sight, unrelated events, where disorder has opposite effects (inhibition or activation)

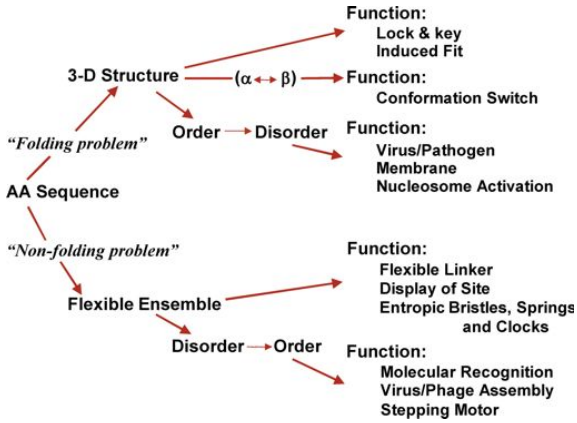


Figure 1.2. A schematic representation of intrinsic disorder in protein function. Adapted from [27]

on different partners or even on the same partner molecule under different conditions. [34] The peculiar properties of IDPs/IDPRs allow them to in play key roles in cellular signaling networks acting as hub proteins. Indeed signaling interactions inside the cell can be considered as "scale-free" networks, which have hubs with many connections, and ends connected just to one neighbor. [35–37] The suitability of IDPs to act as hub proteins is supported by the systematic analysis of known hub proteins and by a several comprehensive bioinformatics studies. [38, 39]

The highly dynamical nature of IDPs in their native state, has also important consequences on the mode of interaction with partners, definitely different from the lock and key mechanism that has for so long influenced our minds. Indeed, IDPs are characterized by very diverse range of binding modes respect to those involving globular proteins. The term "molecular recognition features" (MoRFs) has been introduced to define compact intrinsically disordered interaction interfaces (less than 30 residues) that undergo disorder-to-order transitions upon specific binding into a grove at the surface of the ordered partner. [40] This phenomenon is often referred to as folding upon binding and many examples have appeared in the literature. Structurally, MoRFs are classified according to their fold upon binding: α-MoRFs form α-helices, β-MoRFs form β-strands, and i-MoRFs form structures without a classic secondary structure pattern. [13] Generally, each MoRF contains at least one short linear motif (SLiM), a short sequence (three to ten residues) with the function encoded in the highly evolutionary conserved amino acid pattern. [41,42] However, contrary to the initial belief that folding would be necessary for binding, a variety of different examples showed that some IDPs/IDPRs do not fold upon interaction, but that they rapidly probe the surface of their interaction partner while simultaneously

changing multiple low-affinity binding sites and forming so-called fuzzy complexes. [43,44] Finally, IDPs/IDPRs can form relatively static complexes, which however display a much larger range of interaction modes respect to complexes involving globular proteins. This vast range of interaction modes, described through a variety of terms, such as wrappers, chameleons, penetrators, huggers, intertwined strings, long cylindrical containers, connectors, armature, tweezers and forceps, grabbers, tentacles, pullers, stackers or β-arcs, and MoRFs, is schematically shown in FIgure 1.3. [45]

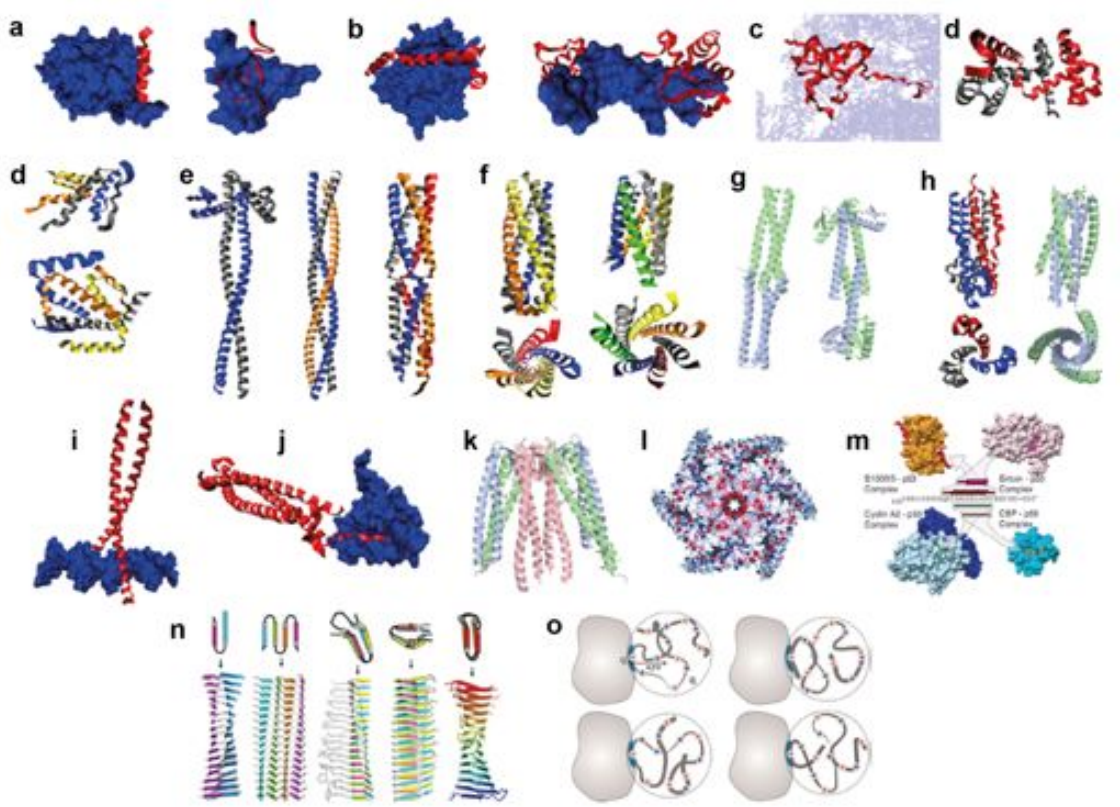


Figure 1.3. An illustrative examples of various interaction modes of intrinsically disordered proteins. **a**. MoRFs; **b**. wrappers; **c**. penetrator; **d**. huggers; **e**. intertwined strings; **f**. long cylindrical containers; **g**. connectors; **h**. armature; **i**. tweezers or forceps; **j**. grabbers; **k**. tentacles; **l**. pullers; **m**. chameleons; **n**. stackers or β-arcs; **o**. fuzzy complexes. Adapted from [45]

1.3 IDPs and diseases

Since disorder is highly abundant in eukaryotes and often plays crucial roles in many biological processes, it is not surprising that IDPs are also implicated in various human diseases.

The most well-known diseases in which they are involved are neurodegenerative diseases, such as Alzheimers and Parkinsons diseases. They are caused by misfolding and aggregation of an IDP, an IDPR or even a short exposed loop that can change conformation, and lead to the formation of amyloid fibrils. [46–48] These pathologies then manifest themselves when significant amounts of filamentous aggregates accumulate in the afflicted tissue, e.g. in the case of the Alzheimers and Parkinsons diseases in brain tissue. [49, 50]

IDPs are also involved in pathologies associated with cell cycle deregulation as many of them are engaged in recognition, regulation, cellular signaling processes. One of such examples is the well-characterized tumor suppressor p53 protein, which occupies the center of a large signaling network. When p53 function is lost, either directly through mutations or indirectly through several other mechanisms, the cell often undergoes cancerous transformation. [51, 52] Many other examples of IDPs involved in cancer, cardiovascular diseases, neurodegenerative diseases, amyloidosis have recently been identified through bioinformatics tools for protein disorder predictions. [53]

IDPs have interesting properties also for viruses. The latter are generally characterized by small genomes, encoding only a few proteins, and need "economical ways" in terms of genome usage to carry out a variety of different functions, interfere with the host cell and hijack cell regulation. One powerful strategy that viruses have developed in order to seize control over the host cell is based on the extensive mimicry of host proteins short linear motifs (SLiMs). [54] Therefore the exposed backbones of IDPs make a larger number of SLiMs accessible in comparison to folded proteins of similar size. The possibility to harbor a variety of different SLiMs in protein sequences is also related to the role of viral IDPs as hub proteins, another interesting property to interfere with the metabolism of the host cell. Many viral IDPs have been recently extensively characterized revealing several interesting mechanisms related to viral infection. [54] Some viruses have the potential to deregulate cellular pathways and thus induce uncontrolled proliferation as an unfortunate side effect of the infection leading to cancer. [55] On the other hand, the viruses with anti-carcinogenic effect were identified as well. [56] The two protein target selected for NMR characterization in this thesis are of viral origin and related to oncogenic processes (E7 from human Papillomavirus and E1A from human Adenovirus). They are described

more in detail in Chapter 4.

1.4 High resolution structural and dynamic investigation of IDPs

The detailed characterization of such diverse structural and dynamical properties of IDPs is not a simple task. X-ray crystallography, historically the main technique to access high-resolution information on large biomolecules, is not suitable for IDP investigation as they rarely crystalize and even if they would, the spatial constrains in the crystal would trap IDP in one of its many conformations. This would provide just information on one of the many states that the protein can adopt, giving a very inaccurate picture of the protein.

Therefore, NMR provides a unique spectroscopic technique to access atomic-resolution information on structurally and dynamically heterogeneous biomolecules. It allows the study of the structural and dynamic properties of IDPs in conditions close to the physiological ones, or directly inside the cell, as well as their interactions with multiple binding partners. [57] For a long time, under the influence of structure-function paradigm, most of the investigated proteins were characterized by a stable 3D structure and guided the optimization of NMR experiments. [58] Thus, new NMR methods need to be designed, or conventional techniques developed for folded proteins need to be adapted to account for the particular spectroscopic properties of IDPs. This topic is discussed in detail in Chapters 2 and 3.

Chapter 2

NMR challenges to study IDPs

A wide range of NMR experiments has been developed throughout the years for the study of globular proteins and their interactions, with the major objective of providing high-resolution structural and dynamic information. These experiments provide the natural starting point for the NMR investigation of IDPs. However, the peculiar properties of IDPs have a strong impact on NMR experiments and spectral parameters. This means that conventional NMR experiments need to be improved for the specific properties of highly disordered proteins in order to study IDPs of increasing size and complexity.

In order to better explain the impact of different extents of local mobility, structural order and disorder on the NMR spectral parameters, the principal features that distinguish a well-folded globular protein from a largely unfolded and highly flexible IDP are illustrated in Figure 2.1.

Hydrophobic interactions and the formation of hydrogen bonds are the main driving forces for the folding process of a polypeptide chain into a 3D structure, and for the stabilization of compact globular folds. The main consequences of a compact structure, that are of importance for the NMR properties of the molecule, are a higher proton density, the exclusion (to a large extent) of water from the interior of the protein, and, to a first rough approximation, the possibility to describe the random motions of a structured protein by an overall rotational correlation time. The high proton density and the large rotational correlation time makes the ^{1}H-^{1}H NOEs the main source of structural information for globular proteins, whereas they are only of little use in IDP research, where the effect is mainly restricted to proton pairs that are close in the primary sequence and is on average smaller than for globular proteins.

The dynamic behavior of IDPs is significantly different from that of structured proteins, as a result of the different energy landscape that is responsible for the high flexibility

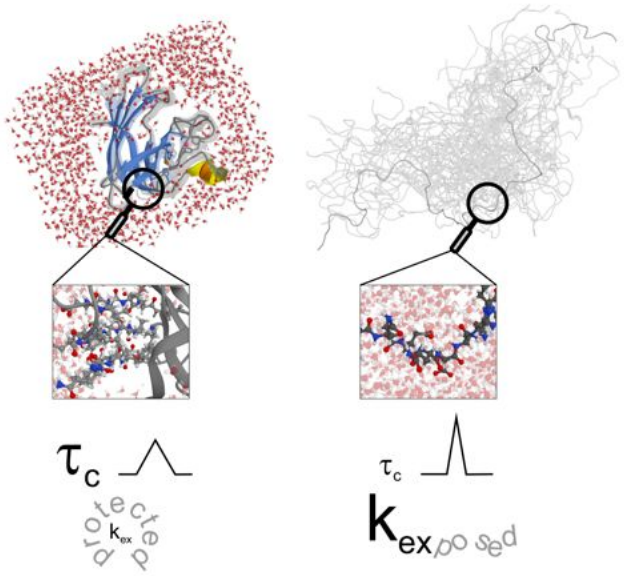


Figure 2.1. Schematic illustration of differences between structured and intrinsically disordered proteins (for clarity reasons the molecules of water were omitted in the ensemble of IDP conformers).

typical of IDPs and allows easy inter-conversion between a many different conformations. Therefore, even to a first approximation, a single overall rotational correlation time cannot be defined for IDPs and local effective correlation times are on average smaller respect to those used to describe structured proteins of similar size. The differences in ^{1}H density and molecular dynamics have also a strong influence on the chemical shifts and the relaxation parameters. Conformational averaging drastically reduces contributions to chemical shifts deriving from the local environment and causes problems of severe resonance overlap. The high flexibility of IDPs has a strong impact on nuclear relaxation rates. Depending on the time scale of motions, several situations can be encountered, ranging from very large contributions to relaxation rates due to exchange processes, up to the case where lines are broadened beyond detection, to very fast motions that instead largely reduce contributions to relaxation rates, as often encountered for highly flexible IDPs amenable to NMR characterization. The latter type of IDPs result thus very good targets for NMR investigations as a variety of complex experiments can be planned and the narrow line-widths contribute to the better resolution of the NMR spectra. Finally, IDPs are also characterized by non-compact, elongated conformations and the exposure of labile protons, e.g. amide protons, to the solvent results in higher exchange rates. Indeed, the measurement of solvent exchange rates is used routinely to distinguish between highly structured and disordered regions of proteins by identifying solvent exposed and solvent

protected amide sites. [59, 60] Approaching physiological pH and temperature, exchange rates of solvent exposed amide protons may broaden resonances beyond detection and in this case alternative nuclear spins respect to amide protons should be detected to access information through NMR.

The amino acid sequences of IDPs also have distinctive features that have an impact on NMR spectra. They can be highly repetitive [19] with long amino acid stretches constituted by only one or by a few amino acid types (low-complexity regions), that are expected to cause further problems of cross peak overlap. In addition, IDPs are typically characterized by large proline content [61] (Figure 2.2), which also needs to be accounted when choosing the most appropriate experimental strategy as amide protons are lacking in prolines which means that they cannot be monitored through 2D ^{1}H-^{15}N correlation experiments. The main bottlenecks for NMR studies of IDPs will be briefly discussed in the following paragraphs.

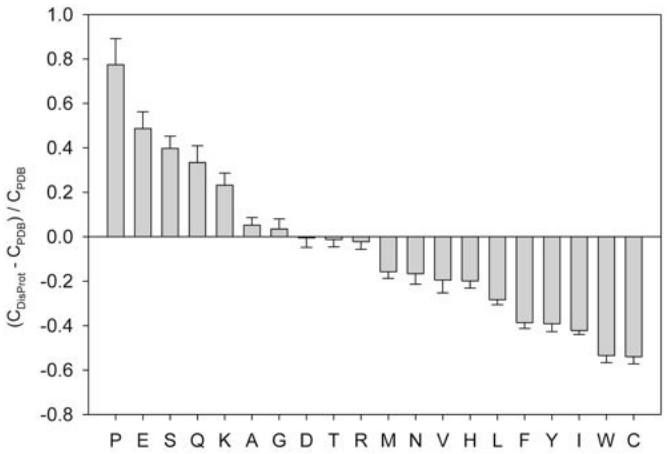


Figure 2.2. Fractional difference in the amino acid composition between the typical IDPs from Disprot database [21] and a set of proteins with the stable 3D structure [62] calculated for each amino acid type. Fractional difference was calculated as: $(C_{DisProt} - C_{PDB})/C_{PDB}$, where $C_{DisProt}$ is the content of a given residue type in a DisPropt database and C_{PDB} is corresponding content in the dataset of well-folded proteins. Positive bars correspond to residues found more abundantly in IDPs, whereas negative bars show residues, in which IDPs are depleted. Adapted from [61]

2.1 Spectral resolution

In order to extract structural and dynamic information for single nuclear sites in an IDP, sufficient spectral resolution to distinguish individual resonances (or correlation peaks) is required. The first consequence of the lack of a stable structure is the averaging of a large

part of the contributions from the local chemical environment to the chemical shift. This results in a drastic reduction of the chemical shift ranges for the different nuclear spin species (^{1}H, ^{13}C, ^{15}N), and thus in strong overlap in the corresponding NMR spectra.

This resolution problem is especially pronounced for ^{1}H NMR. It is well known from studies of unfolded protein states that ^{13}C and ^{15}N nuclei show superior chemical shift dispersion compared to ^{1}H. [63, 64] This is definitely an interesting property provided it also holds for IDPs. As an example, Figure 2.3 shows the chemical shift dispersion observed for three backbone nuclear spins, $^{1}H^{N}$, $^{13}C'$ and ^{15}N in α-synuclein, a well-characterized IDP. It is clear that also in the absence of a global structure the chemical shift dispersion increases from $^{1}H^{N}$ to $^{13}C'$ to ^{15}N. The same holds for aliphatic and aromatic spin pairs, with ^{13}C yielding a higher resolution in the NMR spectrum than the attached ^{1}H. Therefore the improved frequency resolution in ^{13}C and ^{15}N dimensions on multidimensional NMR spectra will be crucial for the study of IDPs.

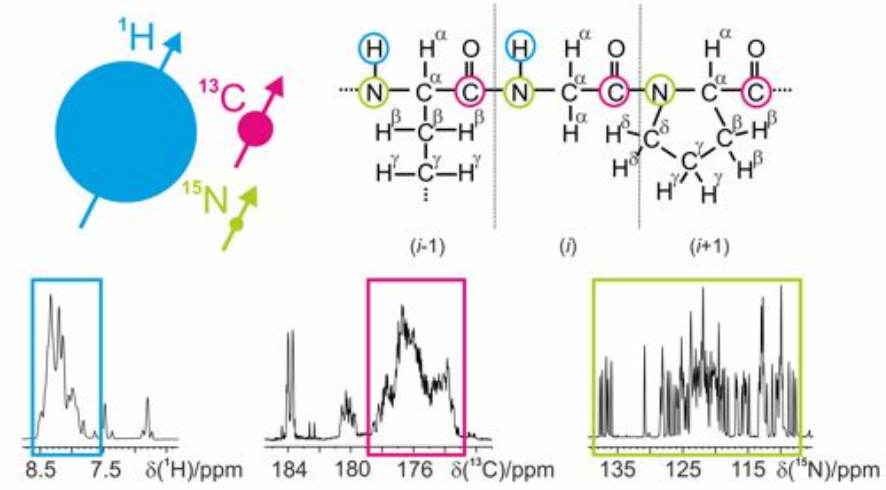


Figure 2.3. The difference of the ^{1}H, ^{13}C and ^{15}N chemical shift dispersion of the IDPs is illustrated on the 1D spectra of α-synuclein. Adapted from [65]

Moreover, because of the lack of structuring chemical environments, the nuclei within the repetitive elements in the primary sequence will resonate at nearly identical NMR frequencies. As a consequence, to obtain resolved NMR correlation peaks for such peptide stretches, more than three dimensions in NMR experiments might be required.

In general, identifying strategies to overcome the problem of resonance/cross-peak overlap is of key importance to be able to study IDPs of increasing size and complexity. Several strategies can be used to address this critical point: exploitation of the favorable

resolution of heteronuclei in NMR spectra; high-dimensional experiments; selection of ^{15}N spin states with favorable transverse relaxation properties; the use of semi-constant time editing in indirect dimension to enhance the spectral resolution while still exploiting spin evolution during a coherence transfer delay; correlation of nuclei of neighboring amino acids; amino acid selection to simplify spectra and identify residue types.

2.2 Experimental sensitivity

From an NMR sensitivity point of view the high flexibility of IDPs has both advantages and disadvantages. The fast timescale molecular motions in IDPs are responsible for reduced effective rotational correlation times, a feature that contributes to long spin coherence lifetimes (long T_2) and narrow NMR lines. In principle, this enables the design of complex high-dimensional NMR pulse schemes to achieve the necessary resolution required for IDPs of large size. It is interesting to note that for highly flexible IDPs the increase in molecular size does not have a major impact on relaxation times and linewidths, while it does for folded proteins. However, for regions with a significant amount of transient structure, this is no longer the case. This results in a large dynamic range of peak intensities observed in the NMR spectra.

This feature becomes even more pronounced for more complex NMR experiments involving an increasing number of transfer steps and frequency editing periods. The effect of partially structured peptide regions and of extensive conformational dynamics on the NMR spectra can become so pronounced that most of the nuclear spins are affected. No solutions, or only very limited ones, remain to study by NMR proteins characterized by strongly line-broadened resonances. Examples that falls into this category are the so-called molten-globule states that represent the “dark side” of biomolecular NMR, in between well folded globular and highly disordered states.

In some cases, sensitivity can also be reduced because of protein aggregation. Tendencies for aggregation can hamper NMR studies of these proteins, because it limits the maximal protein concentration in the NMR sample tube. On the other hand, in some cases higher molecular mass aggregates become amenable to other NMR techniques, such as solid-state MAS NMR. [66–68]

2.3 Experimental time requirements

Sample stability problems are often encountered because of aggregation, as mentioned before, or proteolytic degradation of IDPs. If the temperature stability of the protein permits, the proteases can be deactivated by heat denaturation. However, if the IDP retains significant residual structure, this is often not a valuable option as extensive heat treatment may result in irreversible changes of the structural features. In such cases, the NMR spectra have to be acquired before sample degradation occurs. The requirements of short experimental times and at the same time high spectral resolution, achieved by long acquisition times in all dimensions of a high-dimensional (3D, 4D,...) experiment seem to be contradictory, but they can be reconciled by the use of fast NMR data acquisition techniques such as longitudinal relaxation enhancement [69–71] and non-uniform sampling [72, 73].

2.4 Sample optimization

The conformational dynamics and transient structure of IDPs are highly sensitive to experimental conditions such as buffer composition, salt concentration, pH, or temperature. Under certain conditions, some parts of the IDP may undergo conformational dynamics on a timescale that leads to extensive line broadening in the NMR spectra, and thus missing correlation peaks. Optimization of the sample conditions is thus even more important than for globular proteins.

On the other hand, because of the dependence of the structural features of the IDP on the experimental conditions, one may want to study the IDP under physiological conditions, such as neutral pH and high temperature. Also IDPs are frequently subjected to post-translational modifications, thus to study such events elevated pH and temperature are often required to ensure optimal enzyme activity.

The lack of a stable 3D structure, in combination with increased temperature and neutral pH leads to fast exchange of amide (and other labile) protons with water protons. The chemical exchange rates also depend on the chemical environment (amino acid composition), and they cover a range of two orders of magnitude even for unprotected amide sites. Fast hydrogen exchange leads to extensive line broadening of amide resonances, resulting again in missing correlation peaks and, more generally, a heterogeneous peak intensity distribution.

Finally, *in-cell* NMR can be used in order to experimentally show that IDPs remain

flexible *in vivo* and that disorder is not just an artifact of the chosen sample conditions.

2.5 High content of proline residues

Further problems occur because of the typically large proline content in the amino acid sequence of IDPs as mentioned before. As prolines do not have backbone amide protons, they do not give signals in amide ^{1}H detected NMR spectra and thus present breakpoints in the common sequential backbone resonance assignment strategy based on $^{1}H^{N}$ detected triple-resonance experiments. In addition proline residues cannot be investigated through 2D ^{1}H-^{15}N correlation experiments, nor though their many variants to determine different kinds of observables (^{15}N relaxation rates, ^{1}H-^{15}N residual dipolar couplings, *etc.*)

Therefore, numerous approaches have been proposed to tackle this problem. The strategy exploiting $^{1}H^{N}$ detection is based either on the "bridging" proline residues [74, 75] or on the selective observation of the signals originating from residues preceding of following proline [76, 77]. Nevertheless, when the stretches of several prolines in row occur in the protein sequence, this approach becomes inefficient. Thus strategies employing $^{1}H^{\alpha}$ or ^{13}C detected experiments were proposed. [78–80] and will be discussed in more detail in the following chapter.

2.6 Fingerprint spectra of an IDP

Even in the absence of a sequence-specific resonance assignment, recording simple 2D NMR "fingerprint" spectra of the IDP provides some useful information on the protein and allows to evaluate the overall sample properties and the feasibility of a subsequent high resolution NMR study. Indeed, the number of cross-peaks detected in the spectrum compared to the number of peaks expected from the primary sequence, indicates whether the majority of the peptide sequence is NMR-visible under the chosen sample conditions, or only part of it. The latter case occurs quite often as many proteins exhibit a significantly heterogeneous nature in terms of structural and dynamic properties leading to conformational exchange and/or aggregation induced line broadening. The observed dispersion of chemical shifts allows identifying whether the protein is structured, partly structured, or highly unstructured.

For such a fingerprint spectrum, a single correlation peak detected per residue, good signal dispersion and a reasonable acquisition time are desired. Nuclear spins at the junction of two adjacent residues are the most suitable for this purpose. The most common

experiments used are therefore ^{1}H detected ^{1}H-^{15}N (HN), and ^{13}C detected ^{13}C-^{15}N (CON) correlated spectra, as shown in Figure 2.4 for human securin, a 200 amino acids IDP with more than 20% proline residues. [81]

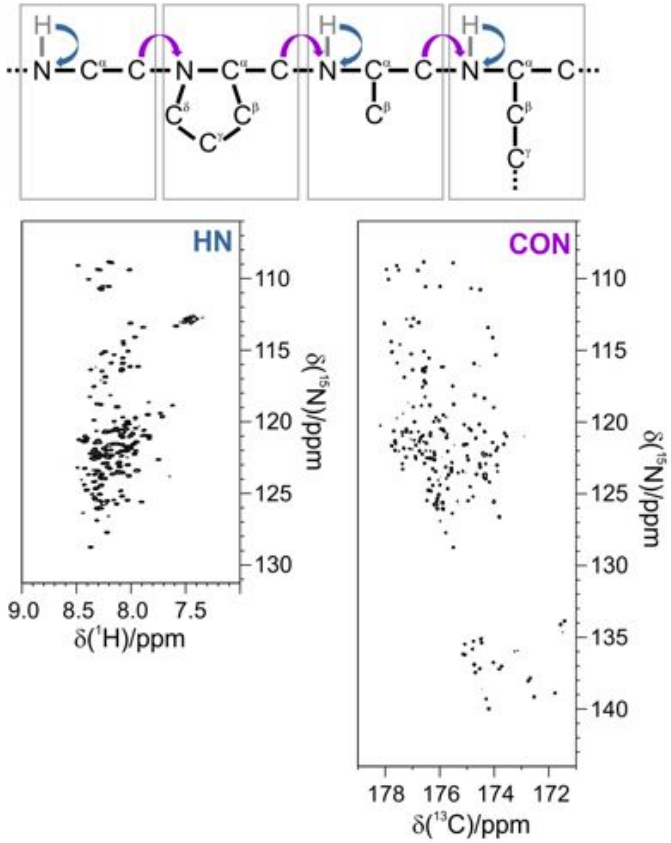


Figure 2.4. The schematic illustration (top) of the correlations observed in the basic 2D ^{1}H-^{15}N HSQC and ^{13}C-^{15}N CON-IPAP experiments. The ^{1}H-^{15}N HSQC (left) and ^{13}C-^{15}N CON-IPAP (right) spectra acquired on the ^{13}C, ^{15}N-labeled sample of the intrinsically disordered human securing. [81] The experiments were acquired at Bruker AVANCE 700 MHz spectrometer equipped with the CPTXO probe.

While the HN experiment is much more sensitive and can be recorded in a significantly shorter time, proline residues are only detected in the CON spectrum that also shows a better spectral resolution. In addition, the CON spectrum does not suffer from hydrogen-exchange-induced line broadening and thus can still be recorded under conditions of high pH and temperature, where many HN peaks are no longer detectable, as shown in Figure 2.5 that reports the HN and CON spectra acquired on α-synuclein with increasing temperature, approaching physiological conditions.

On the other hand, high-quality HN correlation spectra can be recorded on protein

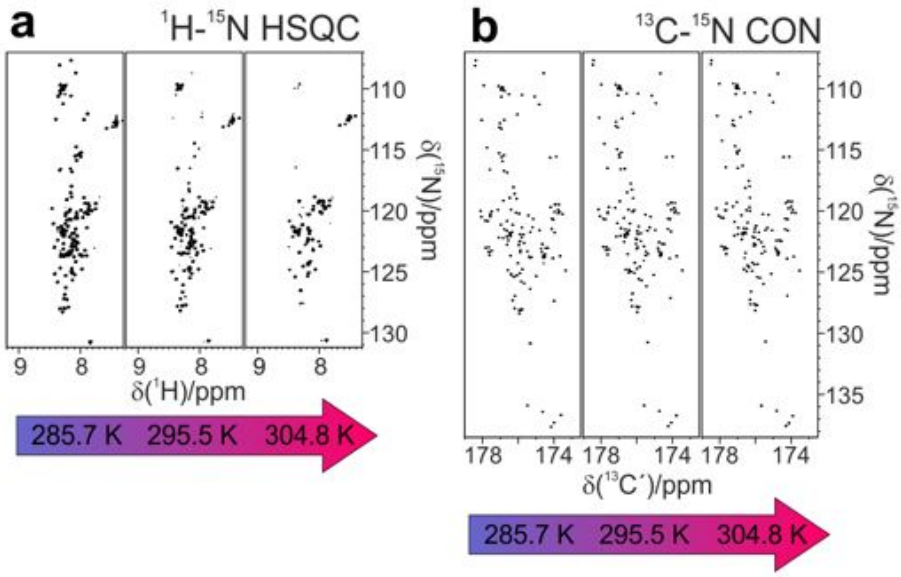


Figure 2.5. The series of 2D HN (**a**) and 2D CON (**b**) spectra acquired on α-synuclein at pH 7.4 at three different temperatures (285.7 K, 295.5 K and 304.8 K).

samples at concentrations of only a few μM on a spectrometer equipped with a cryogenic probe. The 2D HN and CON spectra are thus highly complementary both in terms of detectability and of information content. These 2D correlation experiments can be acquired in a variety of different ways. The most appropriate variants for applications to IDPs are discussed in detail in Chapter 3.

These 2D spectra can also be used to follow changes in the properties of the IDP for changing the experimental conditions such as temperature, pH, ionic strength, buffer, reducing/oxidizing environment, or upon adding potential partners such as metal ions, small molecules, nucleic acid fragments and proteins. They also enable to follow chemical reactions such as the occurrence of post-translational modifications. Finally, they provide an invaluable tool to take a snapshot of a protein inside an entire cell.

Modifications of these basic 2D experiments (HN and CON) enable to determine a variety of observables that report on different properties of the IDP at atomic resolution once sequence-specific assignment becomes available. These include ^{15}N relaxation rates, scalar couplings, residual dipolar couplings (RDCs) [82], paramagnetic relaxation enhancements (PREs) [83] and cross-correlation rates (CCR) [84], as well as solvent exchange rates. All these NMR observables report on the structural and dynamic features of the IDP.

Chapter 3

Methodological aspects

As discussed in the previous chapter the study of IDPs by NMR brings several challenges. Extensive resonance overlap and high exchange rates of the exchangeable protons with the solvent are two major problems, which significantly affect the quality of the spectra, and should be always considered when choosing the set of NMR experiments for investigation of a particular IDP.

During recent years several approaches have been developed in order to exploit favorable properties of IDPs and minimize negative effects originating from the lack of stable 3D structure. Long evolution delays in indirect dimensions of NMR experiments are nowadays routinely used either exploiting real-time or semi-constant time evolution. Proton detected experiments employing transverse relaxation-optimized spectroscopy [85] in combination with the longitudinal relaxation enhancement techniques [86] (BEST-TROSY), have proven to be a powerful tool for IDP characterization, providing well-resolved spectra in relatively short experimental time. [77] These experiments are well suited for low concentrated samples or when time stability of the IDP is the issue. On the other hand, the use of $^{1}H^{N}$ detected experiments limits the range of potential sample condition, since increasing temperature and pH promotes fast hydrogen exchange causing broad lines and sensitivity losses. As mentioned before, proline residues, often abundant in IDPs, are not observable as they lack the amide proton. In order to overcome these problems, the non-exchangeable $^{1}H^{\alpha}$ protons can be exploited. $^{1}H^{\alpha}$ detected experiments have thus been proposed for sequence-specific assignment of IDPs. [79,87] These experiments can be used to characterize IDPs at elevated temperature and pH, and also enable us to detect peaks deriving from proline residues, unlike in the $^{1}H^{N}$ detected experiments. However, the poor chemical shift dispersion when going from amide to alpha proton detection and the signal of the water protons resonating in the middle of the H^{α} region, represent the limiting

factor for the amount of the residue-specific information that can be extracted from the spectra.

Hence, ^{13}C direct detection represents a valuable alternative. Actually, if sample concentration is sufficient, as in many situations of practical interest, ^{13}C detection represents the ideal tool of investigation of IDPs due to a very good compromise between resolution and sensitivity, in particular approaching physiological conditions of temperature and pH. The larger chemical shift dispersion of ^{13}C nuclei compared to that of protons, and intrinsic insensitivity of heteronuclear experiments to the chemical exchange induced line broadening are in favor of using ^{13}C detection for the characterization of IDPs. The ^{13}C direct detected experiments allow to access atomic resolution information even on large IDP studied at elevated temperature and pH, the point where amide proton detected experiments provide only little information on the investigated system. Of course, when using heteronuclear experiments the issues of homonuclear decoupling and sensitivity arise. These aspects are addressed in detail in the following paragraphs.

Naturally, if the IDP sequence contains many repetitive fragments, or the size of IDP reaches certain limit, the 3D experiments are not able to provide sufficient resolution and the use of an additional indirect dimension is necessary. In this case standard sampling scheme would lead to unacceptably long experimental times, thus non-uniform sampling methods are exploited. [72, 73] Strategies for acquisition of multidimensional (>3) experiments, and procedure used to analyze multidimensional spectra, are briefly introduced in the last section of this chapter.

3.1 ^{13}C detected experiments

The first aspect to consider when performing ^{13}C direct detected experiments on uniformly labeled protein samples consists in the presence of large splittings of the ^{13}C resonances in the direct acquisition dimension due to the presence of large homonuclear one-bond couplings, a feature that definitely represents a novelty with respect to what encountered in ^{1}H direct detection. Indeed, despite ^{13}C-^{13}C couplings are very useful in the design of multidimensional NMR experiments, they are responsible for large signals splitting that, as evident from Figure 3.1, need to be suppressed in order to preserve high resolution in the direct acquisition dimension.

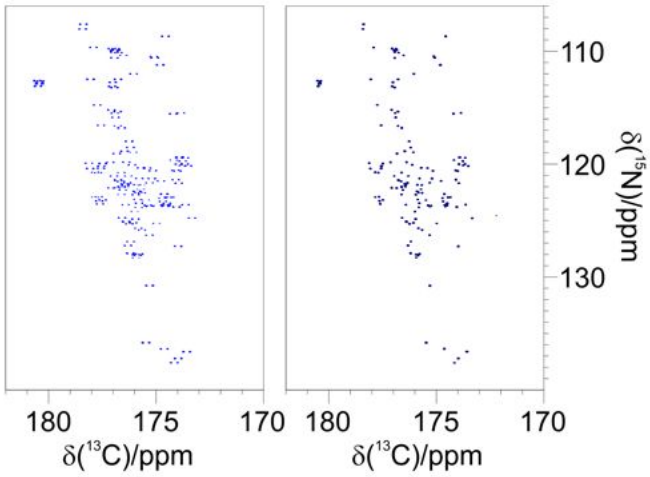


Figure 3.1. Example of 2D ^{13}C detected CON spectra of α-synuclein acquired without and with homonuclear decoupling in direct dimension.

3.1.1 Homonuclear ^{13}C decoupling

To achieve homonuclear ^{13}C decoupling in the direct dimension is more complex than in the indirect dimensions, because it typically requires the application of radiofrequency pulses at a similar frequency than those of the nuclear spins that are in the process of being detected. A possible solution to this problem consists in sharing the detection period between data acquisition and decoupling pulses (band-selective homonuclear decoupling). This, however, reduces the overall sensitivity of the experiment, and introduces Bloch-Siegert phase shifts [88] and decoupling side bands [89] in the spectrum. Recently improved variants exploiting this general idea that however use hard pulses applied at regular intervals during the acquisition period have been proposed. [90] Alternatively, post-acquisition methods, such as data deconvolution using maximum entropy reconstruction, may be used. [91] A last, and arguably the elegant class of methods for ^{13}C homonuclear decoupling uses spin-state selection, also known as "virtual decoupling". [92]

Virtual J_{CC} decoupling uses an spin evolution delay prior to detection, and the recording of, at least, two experiments with different parameter settings. In the first experiment, spin evolution under the J_{CC} coupling is suppressed resulting in an in-phase (IP) line splitting in the detected spectrum, while in the second experiment the J_{CC} coupling evolves for a time $1/(2\ J_{CC})$ resulting in anti-phase (AP) line splitting in the final spectrum. A single resonance line (without splitting) is then obtained by calculating the sum and difference of the two recorded spectra. Finally, the two resulting (sum and difference) spectra are

shifted by an amount $J_{CC}/2$ with respect to each other and added up to yield a single line at the correct chemical shift position. [93, 94] This approach, illustrated in Figure 3.2 for virtual decoupling of $^{13}C^{\alpha}$ in the NMR spectrum of carbonyls, is at the basis of most of the ^{13}C detected multidimensional experiments used for the characterization of IDPs. [92]

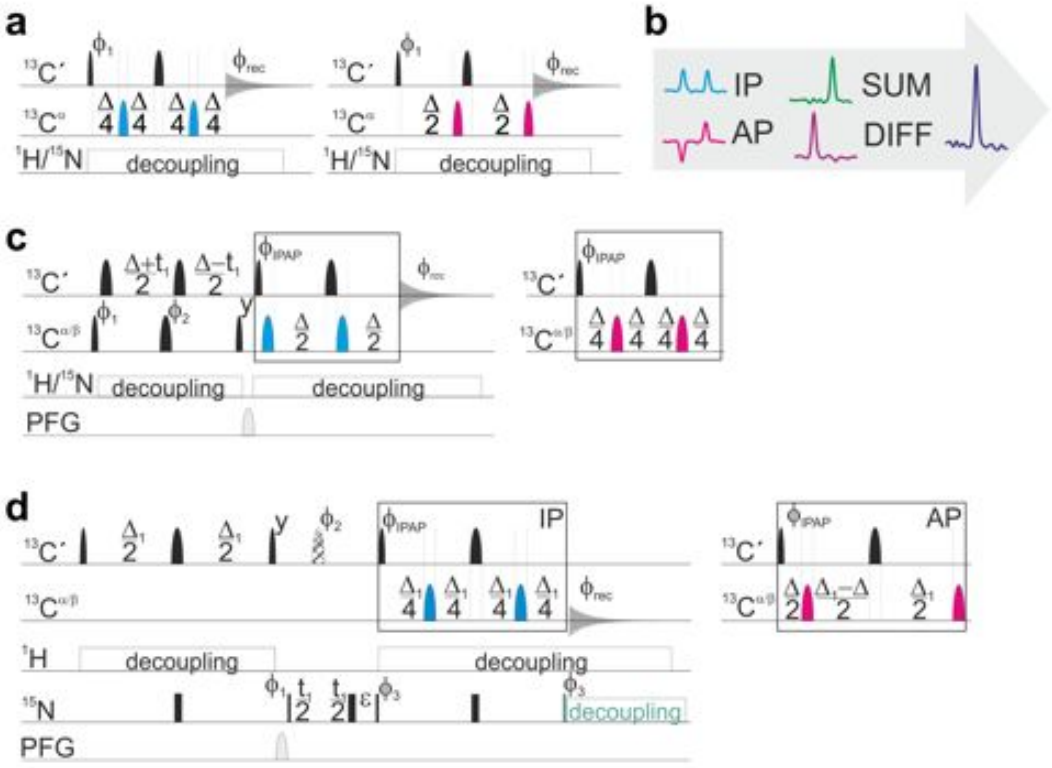


Figure 3.2. ^{13}C detected 1D and 2D experiments with implemented IPAP decoupling sequence. Implementation of $^{13}C'$-$^{13}C^{\alpha}$ IPAP virtual decoupling building blocks in the 1D mode (**a**) and 2D mode (**c**, **d**). Band-selective ^{13}C pulses are denoted by rounded rectangles (narrow and wide ones represent 90° and 180° pulses, respectively). The pulses are applied along the x-axis unless noted differently. **b** The schematic illustration of the post-acquisition processing for obtaining decoupled spectra. [92]

In order to work properly, virtual J_{CC} decoupling requires quite uniform J_{CC} couplings in the protein, which is the case of the large one-bond coupling between alpha carbons and carbonyls ($^1J_{C'C\alpha} \approx 53$ Hz). It is worth noting that the two experiments rely on the same number of pulses and lengths of delays, in order to ensure identical signal perturbation due to pulse imperfections and spin relaxation effects. Another prerequisite of this technique is that the two nuclear spins, e.g. $^{13}C'$ and $^{13}C^{\alpha}$ are sufficiently well separated to allow their selective manipulation through band-selective pulses. There is also a price to pay in

terms of sensitivity for the virtual decoupling method, as during the linear combinations necessary to achieve homonuclear decoupling the thermal noise is increased, thus leading to a reduction of the signal-to-noise ratio.

Virtual decoupling can be included in any pulse sequence ending with in-phase ^{13}C' transverse coherence. Most of the triple resonance experiments based on ^{13}C' direct detection end with a coherence transfer step that involves refocusing of either ^{13}C'-^{15}N or ^{13}C'-^{13}C$^{\alpha}$ anti-phase coherence (or both). Therefore, virtual decoupling can be implemented by slightly changing the position of the ^{13}C$^{\alpha}$ inversion pulses in the last coherence transfer steps, as illustrated in Figure 3.2.

Finally, it is worth noting that as these spin state selective approaches to achieve homonuclear decoupling perform so well, they can also be implemented for heteronuclear decoupling. Indeed, a variant to achieve heteronuclear ^{15}N decoupling has been proposed for both, ^{1}H^{N} and ^{13}C' direct detection experiments. [95, 96] This can be useful when the ^{13}C' coherence lifetimes allow for long acquisition times and/or to reduce the radiofrequency load/heating on the ^{15}N channel during acquisition, which results particularly useful in the fast pulsing regime. [97]

3.2 Improving sensitivity of ^{13}C detected experiments

The sensitivity of NMR experiments is defined as the signal-to-noise ratio (SNR) obtained in the fixed amount of time and is given by the following equation [98]:

$$SNR \propto C \cdot N \cdot B_0^{3/2} \cdot \gamma_{exc} \cdot \gamma_{det}^{3/2} \cdot n_{scan}^{1/2} \tag{3.1}$$

where C is the constant reflecting thermal noise, probe design and electromagnetic properties of the sample, N is the number of nuclei in the active volume of the sample, B_0 the static magnetic field, γ_{exc} the gyromagnetic ratio of the excited nuclei and γ_{det} that of the detected nuclei, and n scan the number of scans.

General conclusions for improving the sensitivity of an NMR experiments derived from equation 3.1 can be divided into two groups depending to what extent they can be influenced by common user. The first group is related to the NMR hardware (magnetic field strength and probe design) and sample conditions (salt content, protein concentration, *etc.*), which are determined by the available NMR instruments in the laboratory and by inherent properties of the studied system. The second group is related to the properties of the molecule under investigation (size, folded/unfolded, local mobility, *etc.*) and to the design of the NMR experiments, where improvements in sensitivity can be achieved by

appropriate choice of the polarization source and by employing techniques, which allow increasing number of scans in the given amount of time, as will be discussed in the following paragraphs. The other important aspect driving the selection of the most appropriate experiments strategy to be adopted is the resolution. Therefore optimal experiments for the study of IDPs are always the result of compromises between optimal sensitivity and optimal resolution.

3.2.1 Starting polarization source

The ^{13}C detected experiments can be designed in the ^{13}Cstart versions, which befits from the fact that carbon atoms are not directly involved in chemical exchange processes and thus effect of exchange line broadening is minimal. The high flexibility of IDPs result in the short T_1 values (of the order of about 1 s), so that not too long recovery delays can be employed. In addition, the ^{1}H-^{13}C NOE effect can be exploited to enhance SNR just by irradiating protons during the recovery delay. Therefore, in many cases of practical interest the sensitivity of ^{13}C start-^{13}C detected versions of experiments, in particular for the simplest 2D experiments such as the CON, CACO and CBCACO, is sufficient to obtain good spectra.

The sensitivity of the ^{13}C detected experiments can be further improved by using protons as starting polarization source. The large one-bond $^{1}J_{CH}$ scalar couplings can be easily used to transfer polarization from ^{1}H to the directly bound heteronuclei while still keeping experiments "*exclusively heteronuclear*". This means that only heteronuclear chemical shifts are frequency labeled in all the spectral dimensions to take advantage of their larger chemical shift dispersion compared to that of ^{1}H, but still exploiting the larger ^{1}H polarization. [96]

However, the use of $^{1}H^{N}$ as starting polarization source renders the experiment susceptible to the loss of information about proline residues and reintroduces a dependence on exchange rates of amide protons with the solvent that impacts on the experimental sensitivity. An alternative solution consists in exploiting $^{1}H^{\alpha}$ as starting polarization source. In this way, all backbone sites can be sampled (including prolines) and exchange processes with the solvent are avoided. [71]

With protons exploited to increase the experimental sensitivity, it is worth thinking about the possibility to implement ^{1}H longitudinal relaxation enhancement (LRE) techniques to reduce the duration of NMR experiments or to increase the sensitivity per unit of time or to increase the resolution. [71, 97]

3.2.2 Longitudinal relaxation enhancement

Usually, a large part of the experimental time in NMR is consumed by the inter-scan delay during which the system is allowed to restore sufficient spin polarization to restart the experiment for a subsequent scan. Its duration depends on the longitudinal relaxation processes. Thus the dependency of the experimental sensitivity on the number of scans can be approximately expressed in the terms of longitudinal relaxation, provided this can be described by a single exponential function with T_1 as the longitudinal relaxation time:

$$SNR_{max} \propto \frac{1}{\sqrt{T_1}} \tag{3.2}$$

Thus, enhancing the longitudinal relaxation efficiency of the excited spins provides a way for increasing experimental sensitivity, and additionally reduces the minimal data acquisition time.

The focus is given to the ^{1}H excitation experiments as they are most widely used. In order to understand the experimental schemes that have been proposed for longitudinal proton relaxation enhancement, the spin interactions that govern proton relaxation are briefly discussed.

There are essentially two main relaxation mechanisms that are responsible for proton longitudinal relaxation in a protein: (i) ^{1}H-^{1}H dipolar interactions; and (ii) hydrogen exchange processes between labile protein protons, e.g. amide and hydroxyl protons, and water solvent protons. The time evolution of the polarization of each proton spin in the molecule is given by the Bloch-McConnell equations [99, 100]:

$$-\frac{d}{dt}\begin{pmatrix} W_{1z}-W_{1z}^0 \\ H_{1z}-H_{1z}^0 \\ H_{2z}-H_{2z}^0 \\ \vdots \\ H_{nz}-H_{nz}^0 \end{pmatrix} = \begin{pmatrix} \rho_W & 0 & 0 & \cdots & 0 \\ 0 & \Sigma_j\rho_{1j} & \sigma_{12} & \cdots & \sigma_{1n} \\ 0 & \sigma_{21} & \Sigma_j\rho_{2j} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \sigma_{n1} & \sigma_{n2} & \cdots & \Sigma_j\rho_{1nj} \end{pmatrix} \begin{pmatrix} W_{1z}-W_{1z}^0 \\ H_{1z}-H_{1z}^0 \\ H_{2z}-H_{2z}^0 \\ \vdots \\ H_{nz}-H_{nz}^0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ -k_{ex,1} & k_{ex,1} & 0 & \cdots & 0 \\ -k_{ex,2} & 0 & k_{ex,2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -k_{ex,n} & 0 & 0 & \cdots & k_{ex,n} \end{pmatrix} \begin{pmatrix} W_{1z} \\ H_{1z} \\ H_{2z} \\ \vdots \\ H_{nz} \end{pmatrix} \tag{3.3}$$

where H_{iz} denotes the z-component of the polarization of proton i and H^0_{iz} is its thermal equilibrium value. The different ρ and σ terms stand for auto- and cross-relaxation rate constants, with values depending on the distance separating the two protons involved as well as the global and local dynamics of the protein experienced at the sites of the two protons. W_z stands for the bulk water polarization, and $k_{ex,i}$ are the hydrogen exchange rates for individual protons with the water.

Equation 3.3 indicates that the relaxation of an individual proton spin depends on the spin state of all other protons in the protein, as well as the bulk water, at the start of the recovery time. The selective spin manipulation of a subset of protons, e.g. amides, while leaving all other protein and water proton spins unperturbed, thus provides an efficient spectroscopic tool for enhancing longitudinal proton spin relaxation.

The contributions of solvent exchange and dipolar cross-relaxation of amide protons depend on the size and residual structure of the IDP, and the samples conditions. Figure 3.3 shows inversion recovery curves of amide protons measured for α-synuclein sample upon selective or non-selective inversion of different sets of proton spins. At low temperature and pH, the dipolar interactions contribute to the effective longitudinal relaxation, while this situation changes as solvent exchange becomes more efficient, at higher temperature and pH.

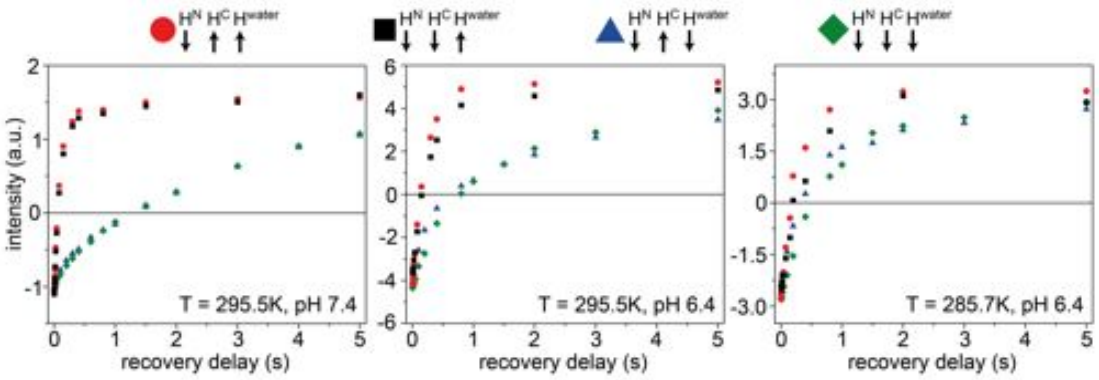


Figure 3.3. Inversion recovery profiles of $^1H^N$ acquired with the variant of $H^{N\text{-flip}}CON$, resp. $H^{\alpha-\text{flip}}CON$, under different initial conditions. Inversion recovery profiles of Alanine 27 of α-synuclein are shown. Different initial conditions of $^1H^N$ inversion recovery profiles are indicated with the following symbols: (●) H^N-selective, (■) H^N-H^C-selective, (▲) H^N-H^{water}-selective and (◆) non-selective. The pH of the sample, and temperature at which experiments were conducted are specified in each graph.

When dipolar interactions provide the dominant relaxation mechanism, the selective spin manipulation allows to reduce effective $^1H^N$ T_1 by a factor three to four. [101, 102] This difference becomes even more pronounced in the case of fast hydrogen exchange.

Under these conditions, the non-selective $^1H^N$ T_1 approach the T_1 of bulk water ($\sim$ 3 s at 25°C), while the selective T_1 is roughly reduced by a factor of 38, resulting in a potential sensitivity gain of a factor of six, according to equation 3.2. This clearly motivates the use of pulse schemes that leave protons that are not exploited as polarization source unperturbed throughout the NMR experiment.

There are several way how to selectively manipulate different kind of proton spins, and thus achieve longitudinal relaxation enhancement. One possibility is the use of band-selective ^1H pulses, especially in the case of amide protons resonating in a frequency range that is well separated from aliphatic and water protons. Pulse sequences exploiting this concept have been termed Band-selective Excitation Short-Transient (BEST) experiments. [69,70] This approach has been implemented for all basic pulse sequence elements (building blocks), e.g. INEPT, sensitivity-enhanced reverse-INEPT (SE-RINEPT or planar mixing), and TROSY (or double-S3CT) required for setting up most of the common triple-resonance experiments. [69, 70, 77, 103]

However, $^1H^\alpha$ are not easy to be irradiated selectively from the other protons through band-selective pulses. A solution consists in exploiting the one-bond scalar coupling ($^1J_{C\alpha H\alpha}$) with the attached carbon ($^{13}C^\alpha$) to selectively manipulate $^1H^\alpha$ protons and flip-back all other protons that are not actively used in the magnetization transfer pathway (H^{flip} approach). [71] Hence by a simple modification of the initial INEPT block in any $^1H^\alpha$-start pulse sequence, the longitudinal relaxation enhancement can be achieved. Of course this approach can be as well implemented for amide protons ($H^{N\text{-}flip}$). [71]

3.3 Multidimensional experiments and non-uniform sampling

As previously mentioned the large signal overlap is one of the major problems when IDPs are investigated through NMR. Multidimensional experiments then represent a utile method to increase resolution. Typically, the amplitude of the NMR signal in each dimension of a multidimensional experiment is measured (sampled) at discrete, uniformly spaced time points. Uniform sampling is required for data processing using the fast Fourier transform (FFT). Hence, the sampled data points form a Cartesian grid with the spacing between time points in each dimension given by the Nyquist theorem. As a result of this uniform data sampling procedure, the experimental time requirement increases by about two orders of magnitude for each additional dimension, making four- and higher-dimensional experiments impractical, or requiring strong compromise for the spectral res-

olution in the indirect dimensions.

In order to overcome time limitations, alternative sampling strategies combined with appropriate processing tools have been developed over the past two decades. [104] These sparse or non-uniform sampling (NUS) techniques have in common significant reduction of the overall number of sampled time points by recording only a subset of the data points of the Cartesian grid. Examples of alternative sampling schemes are shown in Figure 3.4.

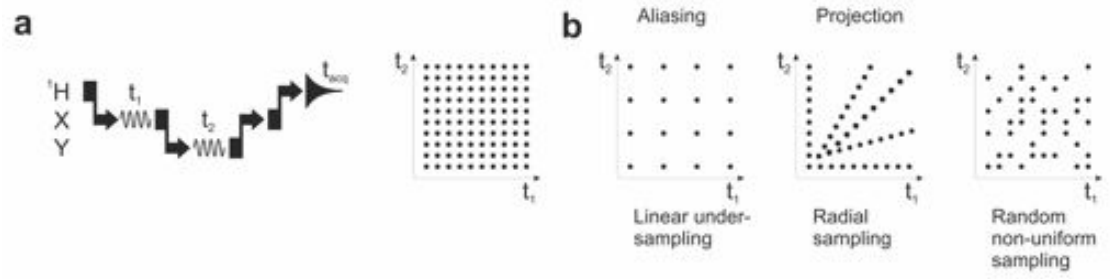


Figure 3.4. Examples of different sampling schemes to speed up the acquisition of a 3D experiment. **a** Schematic illustration of a typical three-dimensional H-X-Y correlation experiment and the conventional time-domain sampling grid required to build the two indirect dimensions. Each point on the grid corresponds to a single repetition of the basic pulse sequence. **b** Some NUS patterns to reduce the experimental time are shown. The linear under sampling is used to fold chemical shifts in an advantageous manner, whereas radial sampling is generally employed in Automated Projection Spectroscopy (APSY) [105]. Instead, random non-uniform sampling schemes are most commonly used with non-linear methods for spectral reconstruction such as MDD [106, 107] or MFT [108]. Adapted from [109]

Some of these sampling grids, e.g. linear under sampling or radial sampling, still yield data sets that can be processed using FFT. [110–112] However, general NUS scheme, where a certain percentage of the grid points is chosen randomly requires alternative, non-linear processing tools. Currently, several non-linear algorithms for processing non-uniformly sampled data are available and well-established, such as multidimensional decomposition (MDD) [106,107], compressed sensing (CS) [113,114], multidimensional Fourier transform (MFT) [108]. The common feature of all these methods is that they aim to find the NMR spectrum that, when applying inverse Fourier transform, best reproduces the measured time data points. As in the case of sparse data sampling, this is an underdetermined computational problem. Therefore, some additional assumptions have to be made to choose the most likely spectrum out of the manifold of possible solutions. MDD exploits the prior knowledge that NMR signals are the direct product of Lorentzian (or Gaussian) line shapes in each frequency dimension. CS performs an $l1$-norm minimization to recover the sparsest spectrum. Both methods have been implemented in the lasts versions of the

Bruker TopSpin software and are particularly useful for processing of 2D and 3D spectra acquired in NUS mode. For processing of 4D and 5D spectra the MFT, which performs discrete multidimensional Fourier transform of the data using appropriate data weighting and filtering for the reduction of sampling noise, has proven to give better results.

3.3.1 Analysis and inspection of multidimensional spectra

One of the main barriers to the use and diffusion of high dimensional NMR experiments has always been the idea that more than three dimensions could be conceptually very difficult to handle. Indeed, while it is easy to visualize three-dimensional objects, the same does not hold for objects characterized by a higher number of dimensions. A possible solution consists in breaking them down into combinations of objects of lower dimensionality, as also done when inspecting three-dimensional spectra. Indeed most of the available software applications for the analysis of 3D NMR spectra display only two dimensions at the time, at a specific frequency. In a similar way, higher dimensional NMR experiments can be visualized by inspecting two dimensions at a time, which are instead associated to more than one frequency (for example a peak in a 2D or 3D reference spectrum). This is possible, as in most cases, higher dimensional experiments are extensions of 2D or 3D spectra in which additional information is encoded in the additional indirect dimensions. This idea is at the basis of the approach known as Sparse Multidimensional Fourier Transform (SMFT) [115], proposed to process spectra of higher dimensionality than three and help in the visualization and analysis of multidimensional spectra.

SMFT proposes an innovative strategy to manually examine four- or five-dimensional spectra in a very simple and intuitive way. NUS data are processed with the MFT algorithm, but only at some arbitrarily chosen frequencies (hence the prefix "sparse"): in this way, the inspection of the full spectrum is reduced to inspection of a limited number of spectral regions, a small subset with respect to the n-dimensional space of the full spectrum. For convenience, these cross-sections are processed as two-dimensional spectra. Therefore, four- and five-dimensional spectra appear as a series of 2D spectra to which two or three further frequencies are associated, respectively. Of course, a prior knowledge of these frequencies is needed. Hence, together with the 4D or 5D experiment, respectively a 2D or 3D spectrum, called "basis spectrum", which shares the same correlations with the higher dimensional one, should be acquired to retrieve them. In this way, the dimensions in which the frequencies of the peaks are already known are not processed, whereas the additional two dimensions containing the new information are fully computed, as the position of the signals are unknown there. So, for example, a 4D spectrum in which δ_i,

δ_j, δ_l and δ_m chemical shift evolutions are frequency labelled is transformed as a series of 2D cross-sections (e.g. δ_i and δ_j), each one extracted at a specific pair of the other two frequencies (δ_l and δ_m) collected in the basis spectrum. Analogously, a 5D spectrum can be conceptually split into a 3D basis spectrum in which each peak is then coupled to a further 2D spectrum, containing the new information with respect to the basis spectrum. This approach is schematically illustrated in Figure 3.5.

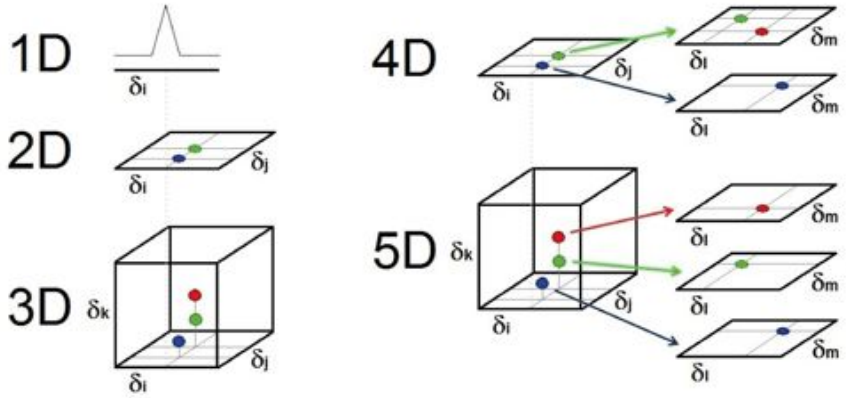


Figure 3.5. A schematic illustration of a possible methods how to visualize multidimensional NMR spectra. Spectra up to three dimensions are visualized in the standard way, whereas 4D and 5D spectra are presented as processed by SMFT algorithm. A 4D spectrum is imagined as a 2D basis spectrum, which shares two dimensions with the 4D spectrum (δ_i and δ_j), in which each peak is associated to another 2D spectrum containing the two further dimensions (δ_l and δ_m). Similarly, a 5D spectrum can be analyzed as a series of two-dimensional spectra (δ_l and δ_m), each one correlated to a given peak of the related 3D basis spectrum (δ_i, δ_j and δ_k). Adapted from [65]

Therefore, SMFT can be successfully exploited to facilitate the analysis of high-dimensionality spectra, since this method provides a way to retrieve all the spectral information content without the need to explore all the vast n-dimensional space of the full spectrum. The visualization of the spectrum through inspection of 2D cross-sections provides also further advantages. For example, it enables the use of automatic peak picking tools, which perform in a very reliable and efficient way, as the signals are well-resolved thanks to the high dimensionality of the experiment. In addition, it allows saving great amounts of disk space, since just a small subset of the full spectrum is actually processed, thus high digital resolution can be set in all the indirect dimensions.

Chapter 4

Viral IDPs

Viruses as obligate intracellular parasites unconditionally rely on the host cellular machinery. Each step of their life cycle is dependent on the high number of diverse interactions with the host signaling and regulatory proteins in order to force the infected cell to create favorable conditions for viral replication. To this goal, viruses need to highjack the host regulatory system, which seems to be a difficult task to achieve with their small genomes.

As discussed in the first chapter, the hub proteins often contain long disordered regions or are even fully disordered. Their function is frequently encoded into brief sequences of amino acids with a specific pattern, called Short Linear Motifs (SLiMs). [41, 42] The evolutionary plasticity of the SLiMs allows easily adding functionality to the protein and thus to quickly rewire cellular pathway, which probably enabled the evolution of the complex, yet highly tuned and robust regulatory systems in higher eukaryotes. [116, 117]

Ironically, this creates an opportunity for the viruses to efficiently alter many regulatory processes by simply mimicking the host motifs. Since SLiMs generally consist of 3 to 10 residues in length, only one or two mutations are sufficient to form a novel motif. In combination with the high mutation rates of virus genome and fixation of advantageous motifs, the hosts SLiMs mimicry represents a very fruitful viral strategy of hijacking the cellular machinery of the infected cell (Figure 4.1). The comprehensive analysis of the highly studied viruses responsible for severe human diseases (human Papillomavirus, Epstein-Barr virus, human T-cell lymphotropic virus, adenovirus, HIV and influenza) revealed that 52 out of ∼150 SLiMs in Eukaryotic Linear Motif database [42] are mimicked by investigated viral systems. [54]

Many viruses use mimicry, for example, to rewire and deregulate signaling pathways, to block apoptosis, to drive cellular proliferation, to highjack the ubiquitination system in order to promote degradation of anti-viral and delay degradation of pro-viral proteins, and

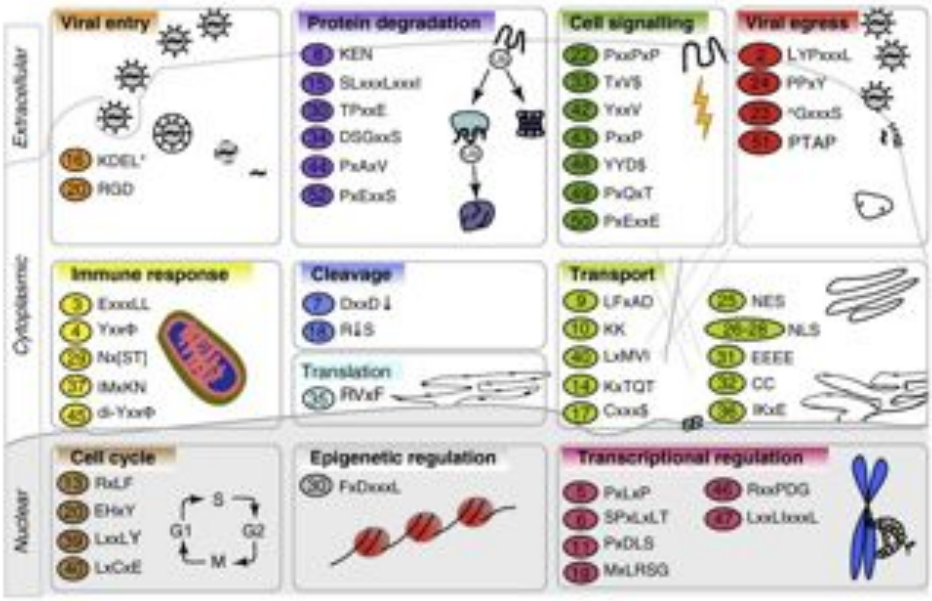


Figure 4.1. Examples of the SLiMs, which are mimicked by various viruses in order to successfully infect the host cell. The motifs are roughly classified by their functions. Adopted from [54]

to deregulate tightly controlled cellular processes, e.g. interaction with the pRB protein releasing E2F transcription factor in order to promote cell cycle to the S phase and set optimal conditions for viral replication. [118]

Unfortunately, some viruses have the potential to trigger uncontrolled proliferation as side effect of altering the cellular signaling network, which then ultimately leads to the tumorigenesis. [119] Generally, only specific subset of the viral hub proteins are responsible for cancer development. Hence, the high resolution characterization of these proteins can help understanding the mechanisms standing behind carcinogenesis. Two IDPs that were identified to play crucial roles in carcinogenesis, and yet not fully characterized on atomic level resolution were chosen as biomedically relevant targets for our study.

4.1 Human Papillomavirus E7 protein

Human Papillomaviruses (HPVs) are small icosahedral non-enveloped viruses with circular double-stranded DNA genome, which are involved in the infection of skin and mucous membrane keratinocytes (the predominant cell type in the epidermis). Infection by majority of the HPV types doesnt cause any physical symptoms and passes without noticing, however in some cases HPV infection drives the development of benign warts or different

kinds of cancer. The sexually transmitted HPV infection has been identified as a causative agent responsible for more than 95% of cervical cancer cases, the second most common womens cancer worldwide. Not least, HPV were also detected in anal, vaginal, vulvar, and penile carcinomas. [120, 121]

Recently, a large number of studies that associate HPV infection also with the squamous cell carcinoma in the head and neck region have been published. [122, 123] However the direct connection between HPV infection and oropharyngeal cancer remains controversial. [124] Overall, it has been estimated that around 10% of cancers are associated with the HPV infection. [121]

More than 170 types of HPV have been identified till now.[1] The HPV types are classified based on their capacity to immortalize human keratinocytes. The types frequently present in cervical and anogenital cancers are called as "high risk" (HR), the types that havent been found in the human tumors are referred as "low risk" (LR). [121] The most prominent member of the HR group is types 16, which is found approximately in 57% of patients with the invasive cervical cancer. [125, 126]

The entire genome of HPV 16 consists of 7.9 kilo base pairs (kbp), encoding six nonstructural (early proteins E1, E2, E4, E5, E6 and E7) and two structural proteins (late proteins L1 and L2) constituting the viral capsid. The E1 protein possesses helicase activity and is essential for papillomavirus replication. [127] The interaction with E2 protein facilitates E1 binding to the origin of replication and thus stimulates viral DNA replication. [128] Additionally, E2 acts also as transcription factor that positively or negatively regulates expression of early genes. [129] The E4, most highly expressed HPV protein, plays a significant role in promoting viral genome amplification and S phase maintenance during differentiation. [130] The E5 protein modulates viral amplification and gene expression in the late phases of the viral life cycle. [131]

The E6 and E7 proteins are crucial for cellular growth regulation, control of the immune system, and progression to a cancerous state. Coordinate action of both proteins is necessary to the development of an invasive carcinoma. [132] However, they are also able to immortalize infected cells independently. [133, 134] The E6 protein promotes carcinogenesis by binding to the p53 tumor suppressor and stimulating its degradation by cellular ubiquitin proteolysis system [135], while E7 interacts with retinoblastoma tumor suppressor protein (pRb) [136], an important regulator of cell growth. Each of the early HPV proteins interacts with a high variety of cellular partners, their interactome was recently reviewed by Xue et al. [137] Despite of the importance of both proteins, E6 and E7, only

[1] http://pave.niaid.nih.gov/

E6 has been fully characterized by high resolution techniques.

E7 is a small protein (98 residues in the HPV 16 variant) constituted by 3 conserved regions (CRs), which share substantial amino acid sequence similarities throughout different HPV types. [138] The CR1 and CR2 localized in the N-terminal region exhibit high propensity to be disordered. The CR3, localized in the C-terminal region, contains two zinc binding CxxC motifs separated by 29 amino acids [139], and is expected to be less disordered. [140] High resolution structural studies have been done solely on the isolated CR3 region, the variant from LR HPV 1a was characterized by X-ray crystallography [141] and HR HVP 45 variant by NMR [142]. However, no 3D structure is available for the HR HPV 16 variant. In addition, the full length protein has never been characterized by high resolution techniques, despite its importance for the understanding of the molecular basis of the many interactions in which it is engaged (Figure 4.2).

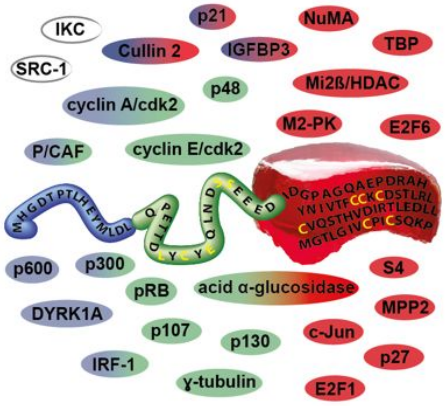


Figure 4.2. Schematic illustration of the human Papillomavirus 16 E7 protein including the amino acid sequence and its division into the three conserved regions (CR1 is in blue, CR2 in green and CR3 in red). In order to present the idea of the complex interaction network of E7, the subset of its cellular targets are shown. The color-coding reflects what part of E7 is involved in the interaction with the particular binding partner.

As mentioned before, the most prominent interaction of E7 is the one with the pRb. The structure of the complex between short oligopeptide from the disordered CR2 region containing the LxCxE motif [136] and pRb was determined and provided useful hints on the interaction mechanism. [143] However, it was shown that full-length E7 has significantly higher affinity for pRb respect to the isolated LxCxE fragment [136] indicating that also other parts of E7 are involved in binding to pRb. [144]

Nevertheless, the binding of pRb is not sufficient for cell immortalization [145] and interactions with other cellular proteins is required for malignant transformation. To en-

hance pRB degradation E7 binds to several other partners, such as the cullin 2 ubiquitin ligase complex bound by the CR1 and C-terminal regions of the E7 protein. [146] The interaction of CR3 region with the subunit 4 (S4) ATPase enhances APTase activity of the S26 proteasome, thus providing additional pRb degradation pathway. [147] E7 interact with a wide range of other cellular signaling and regulatory proteins, such as retinoblastoma related tumor suppressors p107 [148] and p130 [149], a cell cycle regulator cyclin-dependent kinase 2 (CDK2), and its inhibitors p21 [150] and p27 [151]. These represent just a few examples of how E7 prevents apoptosis and promote cellular proliferation. The more comprehensive picture of the complex interaction network of E7 is provided in recent review by Roman et al. [152] Hence, it is of high importance to obtain high resolution information on this oncogenic protein in order to elucidate the molecular basis of the complex mechanisms behind HPV associated diseases.

4.2 Human Adenovirus E1A protein

Other viruses exploit similar strategy to the one used by HPV to highjack cellular regulation, and proteins carrying similar features can be found throughout the viral realm. The analogous regions to CR1 and CR2 from HVP E7 can be found in the adenovirus E1A protein. [149]

Human adenoviruses (HAdVs) are large non-enveloped viruses that cause a variety of gastrointestinal, respiratory and eye infection. In addition, recombinant versions of adenovirus have been extensively used as vectors for gene therapy and nowadays are the most frequently used viral vectors in clinical trials. [153] Since the late sixties, when the first HAdV serotypes were identified to promote tumor genesis in hamsters [154], more than 50 other serotypes were found to cause malignant transformation of rodent cells. These findings established HAdV as a model system to study processes leading to the cancer development. Investigation of its viral function has contributed to reveal important aspects of cellular physiology such as the immunological response of the infected cell, cell cycle control, apoptosis and so on. [155] The HAdV 2 was one of the first adenovirus serotype with the completely sequenced genome and since then each of its protein has been extensively investigated, mainly by molecular biology techniques (biological assays, mutation studies, *etc.*). Nowadays HAdV 2 and 5 (high homology between the two genomes [156]) are the most studied serotypes.

The HAdV 2 and 5 genome consists of approximately 36 kbp linear double stranded DNA stored inside of an icosahedral nucleocapsid. The five early (E1A, E1B, E2 early, E3,

E4) and three delayed-early genes (IX, Iva2, E2 late) are successively transcribed early after infection by the host RNA polymerase II and are responsible for creating an optimal environment for viral replication. [157] Each of the early transcripts is alternatively spliced, generating several mRNAs. The capsid components and proteins involved in capsid assembly encoded by L1-L5 transcription units are expressed later in viral reproductive cycle. [158] The products of the E1A gene are the first proteins expressed early after infection and are essential for productive viral infection. They are primarily responsible for the modulation of the host transcription machinery, promoting transcription of succeeding viral genes, and for reprogramming cell cycle regulation, forcing the cell to enter into S phase. [159, 160] They also play important roles in suppressing the inflammatory response and in blocking apoptosis.

The two major proteins of 289 and 243 amino acids encoded by E1A gene are expressed at high levels right after infection. These proteins arise from alternative splicing of the same pre-mRNA and differ only by the presence of a 46 residue long internal region in the larger one. They are referred as E1A-13S and E1A-12S, based on the sedimentation coefficient of their respective mRNAs. The E1A gene encodes three additional proteins of 217 (11S), 171 (10S) and 55 (9S) residues that are expressed in the latter stage of adenovirus infection. All E1A proteins are spliced keeping the same translational reading frame except of 9S, which shares only 28 common residues with the other forms of E1A, but surprisingly is still able to independently activate expression of viral genes and promote HAdV replication. [161]

The primary sequence alignment of E1A-13S protein from several HAdV serotypes revealed four conserved regions (CR1-CR4) connected by less conserved segments (Figure 4.3). [162] Further bioinformatics analyses using disorder predictors suggest that a large portion of E1A has a high propensity to be disorder, and only short region in N-terminal part and CR3 are expected to be structured. [12,163] This prediction is supported by the extensive analysis of high number of mutation studies, which has shown that insertions or deletions generally do not globally affect the activity of E1A, but interfere only with specific subsets of function. [164] The E1A has been reported to alter or inhibit the function of more than 50 distinct cellular proteins. [165] These interactions are mediated through SLiMs, hence the identification of binding motifs throughout the primary sequence is important for the understanding of the function of E1A and of hub proteins more in general. A brief description of the main SLiMs identified in E1A sequence, their interaction partners and their functional role in the context of the cellular interaction network follows bellow.

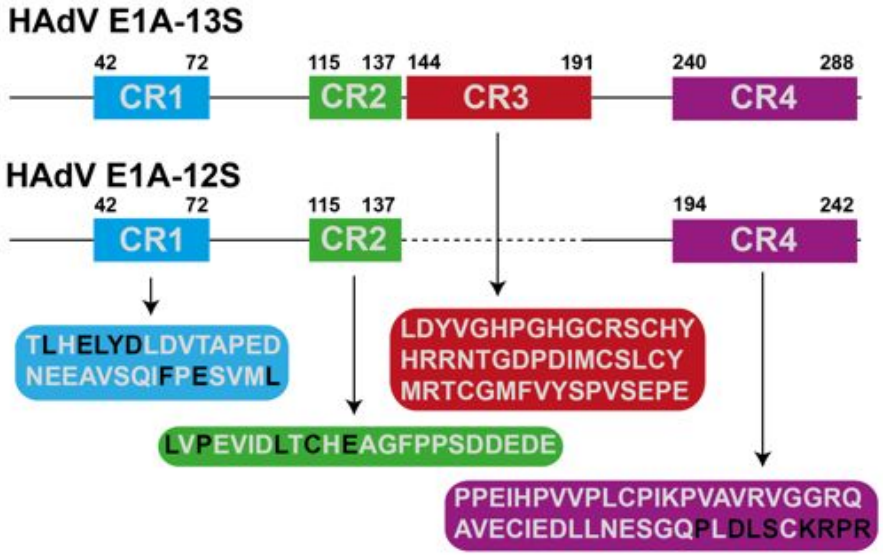


Figure 4.3. Schematic representation of the 13S and 12S forms of E1A protein from HAdV 2, their division into the conserved regions including amino acid sequence. The SLiMs involved in the interactions with the important cellular target are indicated in black.

The first SLiM of the HAdV E1A found in the N-terminal region is LxxLIxxxL (residues 20-28), so called CoRNR box motif, which binds to the unliganded thyroid hormone receptor (TR) and thus deregulate gene transcription. [166] Overall, a large repertoire of more than 15 cellular targets including transcriptional co-activators p300/CBP [167], transformation/transcription domain-associated protein (TRRAP) [168], p400 [169], pCAF [170] and TATA box binding protein (TBP) [171], and the S4 and S8 subunits of the 19S proteasome regulatory component [172], have been reported to date to interact with E1A, reflecting the high functional heterogeneity of this short amino acid region. This region is also predicted to contain a partially populated amphipathic α-helix that contributes to the interaction with some of its many binding partners.

The CR1 cooperates with the N-termini in binding of several transcriptional regulators. It interacts with the p300/CBP via FxDxxxL (residues 66-72) SLiM [173] and with the TRRAP via ELYD (residues 45-48) [168]. The latter motif partially overlaps with the Retinoblastoma tumor suppressor protein binding α-MoRF, LxELY (residues 43-47), which directly competes for the same interaction site as E2F. [174] Additionally, the tight interaction between the pRb and the CR2 region of E1A actually increases the apparent local concentration of pRb, which allows lower-affinity pRb binding motif of CR1 to bind and further stimulates the release of E2F. Free E2Fs then activates transcription of the

early viral genes and genes specific for the cellular S-phase. [174]

The CR2 binds to the pRb through the well-known and well-characterized LxCxE (residues 122-126) SLiM. [175] This was the first interaction identified for the E1A protein. [176] This important motif was subsequently found also in other viral and cellular proteins that also interact with pRb, like the above mentioned E7 from HPV. [136] The mutation studies in the pRb binding motif additionally showed that E1A regions interacting with the small ubiquitin-like moeity (SUMO) conjugase UBC9 [177] and the region interacting with S2 component of the 19S regulatory complex of the 26S proteasome [178] partially overlap with LxCxE SLiM. Another SLiM found in CR2 is the PxLxP motif (residues 113-117) that mediates the interaction with the transcriptional repressor BS69, which however also requires CR3 for successful binding as E1A-12S (missing CR3) exhibit much lower affinity. [179] Through these interactions CR2 interferes with several crucial cellular pathways such as the SUMO pathway that basically regulates all aspects of cell function [180], and is thus essential for a number of E1A functions that are necessary for virus replication like activation of viral early gene expression and stimulation of the infected cell to enter the cell cycle. [164] In addition, CR2 mediates the oncogenic transformation of rodent cells, stabilizes p53, and induce apoptosis. [178, 181–183]

The CR3 region, unique for the E1A-13S form, involved in transcriptional activation contains two CxxC zinc-binding motifs (residues 154-157 and 171-174) creating a minimal zinc-finger subdomain. The fact that almost any deletion mutant cause loss of function [184] and that no SLiMs were identified in this region points to conclusion that CR3 structure is required to interact with its cellular targets such as transcriptional activators TATA binding protein (TBP) and p300/CREB-binding protein (CBP) [185], transcriptional repressors BS69 [179] and GCN5 [186]. Moreover it was shown that not only the N-terminal part, but also the CR3 interacts with the 19S proteasome subunit and thus contributes to its potent transcriptional activation by promoting proteolysis of E1A and other locally associated factors bound to chromatin. [187] These observations indicate that the CR3 region uses a complex mechanism to modulate transcription of the early genes.

In contrary to the previous E1A regions, relatively little is known about the function and binding partners of the CR4 region. There are only a few known proteins that associate with this region. The best described interaction of the CR4 is with the transcriptional co-repressor CtBP, which binds to a PxDLS motif of CR4 (residues 279-283 in E1A-13S or 233-237 in E1A-12S, respectively). This interaction dissociates CtBP from sequence specific transcription factors, inducing repression of cellular genes. [188] The other cellular targets of CR4 are tyrosine-regulated kinase (DYRK) proteins and importin-α. The

DYRK proteins are implicated in regulation of cell survival, proliferation and differentiation, and when bound to the short E1A region that overlaps with CtBP binding site their kinase activity increases. [189] The CR4 interacts with the importin-α through KRPRP motif (residues 285-289 in E1A-13S or 239-243 in E1A-12S, respectively), promoting rapid import of E1A to the cellular nucleus. [190]

Our understanding of the function of E1A is mainly based on biological studies and only very little high resolution information on E1A is available. The atomic resolution study of the short E1A fragment interacting with TAZ2 domain of CBP/p300 complex has helped to better understand of the molecular basis of this interaction. [191] However, it is important to investigate interactions in the context of the full length E1A protein, as it seems likely that it employs a complex system of multiple SLiMs, which bind with low affinity to similar targets, such as p300, in order to form strong and highly specific interactions.

Chapter 5

Results

The scientific context introduced in the previous chapters provided the motivation for my research activities during the doctoral studies. The results are summarized in several scientific publications and in a book chapter presented in the next paragraphs. I would like here to briefly outline the overall general lines of development followed throughout my research work.

The development of new NMR methods for the characterization of IDPs and their application for the study of biologically relevant proteins were the two main lines of my research activities. Needless to say that these two lines of research are highly interconnected, as novel NMR methods are generally inspired by biologically relevant questions, developed on well-known proteins which allow the assessment of their performances, applied to tackle the characterization of challenging proteins, which in turn provides new hints for further developments.

As extensively discussed in the previous chapters, NMR represents a unique spectroscopic tool for the high resolution investigation of IDPs. However the overall general properties of IDPs have strong impact on NMR observables, posing new questions with respect to the study of structured proteins. So it is worth to consider the critical points, evaluate which spin properties are the most appropriate to overcome them and design new NMR methods that allow exploiting at best the unique features of this powerful spectroscopic technique and in this way contribute to the exciting field of research on IDPs.

The major consequences of the lack of a three-dimensional structure and of the pronounced flexibility characteristic of IDPs include extensive resonance overlap and fast exchange processes with the solvent. In this frame ^{13}C detected experiments provides a valuable tool to address the critical points and thus provided the starting point for the development of new experimental variants. The observation that approaching physiological

conditions of pH and temperature 2D CON experiments were able to provide information that cannot be accessed through 2D HN correlation experiments constituted a solid basis in support of the usefulness of the whole suite of ^{13}C detected NMR experiments to study IDPs close to physiological conditions. The observation of such pronounced effects due to exchange processes with the solvent, which represents a general distinctive feature of highly flexible IDPs, motivated further investigation of their impact on different NMR observables to identify weather they could be actively exploited to achieve new information or to improve the existing experiments.

Simple 1D ^{1}H inversion recovery (IR) profiles acquired on different proteins characterized by different structural and dynamic properties (both folded and unfolded/IDP), in different conditions (pH and T) revealed that exchange effects have a strong impact on the IR profiles of exchangeable protons. In particular, by selective manipulation of amide protons in IDPs near physiological conditions amazing longitudinal relaxation enhancement (LRE) effects can be monitored with $^{1}H^{N}$ T_1 passing from values close to the T_1 of water protons (about 3 s) to about 100 ms by selective inversion of amide protons rather than through non-selective inversion. These observations therefore motivated a more accurate study of ^{1}H longitudinal relaxation in IDPs, on the origin of LRE effects and on how these are modulated by the sample conditions and by the local properties of the molecule. The study of ^{1}H longitudinal relaxation properties in IDPs, of general interest for optimal design of NMR experiments, was exploited for the design of additional ^{13}C detected experiments, still keeping them exclusively heteronuclear in order to benefit by contribution of heteronuclear chemical shifts to improving the resolution of NMR spectra of IDPs.

A first application of the impressive LRE effects identified for amide protons under near to physiological conditions consists in exploiting it to reduce the duration of NMR experiments (Paragraph 5.1.1). The BEST approach was thus implemented for the acquisition of 2D CON spectra and this combination revealed interesting general properties (discussed in detail in the publication) and most importantly provided impressive reduction of the minimal duration of an NMR experiment. The potential of this new experiment, the $H^{N\text{-BEST}}$CON, is also demonstrated by applications to the study of IDPs *in-cell* as well as for monitoring the occurrence of post-translational modifications in real time.

Along similar lines, ^{13}C detected NMR experiments were designed to accurately monitor at atomic resolution ^{1}H longitudinal relaxation involving not only $^{1}H^{N}$ but also $^{1}H^{\alpha}$ protons (Paragraph 5.1.2). These experiments, setup and tested on ubiquitin and α-synuclein as representatives of a well-folded and of disordered protein, provide a picture of the major contributions to ^{1}H longitudinal relaxation in IDPs as well as the major

differences respect to folded proteins. These data provide valuable information for the study of local properties of structurally and dynamically heterogeneous proteins such as local compactness, local flexibility and solvent accessibility. In addition, the knowledge of ^{1}H longitudinal relaxation in IDPs also provides a guide to the design of new experiments optimized for the study of IDPs. In particular, while for $^{1}H^{N}$ detected experiments LRE is a useful technique both for IDPs and for folded proteins, it is not useful for $^{1}H^{\alpha}$. On the other hand, $^{1}H^{\alpha}$ are not influenced by exchange processes with the solvent and thus can be studied at higher temperature; the increase in temperature itself causes pronounced enhancement in the longitudinal $^{1}H^{\alpha}$ relaxation without the need of implementing approaches to exploit LRE in pulse sequences. These findings therefore can be of general interest for the selection and design of the most appropriate experimental strategy for specific conditions. For example $^{1}H^{N}$ provide a good probe to investigate IDPs at T and pH values where solvent exchange contributes to LRE but does not excessively contribute to linewidths; ^{13}C and $^{1}H^{\alpha}$ instead offer complementary useful probes to study IDPs at higher T and pH values as they are not sensitive to solvent exchange effects and at higher temperatures benefit by faster longitudinal relaxation and narrower lines.

Therefore these ideas were exploited for the design of additional multidimensional NMR experiments for the sequence specific assignment of IDPs (Paragraph 5.1.3). A suite of multidimensional experiments based on ^{13}C detection was recently developed to contribute to overcoming the problem of extensive overlap in spectra of IDPs of large size. These experiments of higher dimensionality than 3 (4D, 5D), which exploit non-uniform sampling (NUS) of the data and ^{1}H polarization as a starting source, provide an excellent resolution not only in the direct acquisition dimension but also in indirect dimensions of NMR experiments. However, in case of IDPs characterized by limited solubility or short lifetime, ^{13}C detected experiments may not be sufficient for the complete investigation of the protein (on the grounds of the lower gyromagnetic ratio of ^{13}C). Therefore two new experiments, either based on $^{1}H^{N}$ start-$^{1}H^{N}$ detect or $^{1}H^{\alpha}$ start-$^{1}H^{\alpha}$ detect, that exploit the good resolution of C'_{i}-N_{i+1} cross-peaks in indirect dimensions for sequence specific assignment were developed. This can thus be achieved correlating C'_{i}-N_{i+1} cross-peaks to the following (C'_{i+1}-N_{i+2}) or previous (C'_{i-1}-N_{i}) ones. The two experiments, tested on α-synuclein, give optimal result in different experimental conditions (H_2O, low T for the $^{1}H^{N}$-based variant, D_2O, high T for the $^{1}H^{\alpha}$-based variant) and provide an additional contribution to the overall "CON-CON" strategy for the sequential assignment of IDPs combining experiments based on $^{13}C'$, $^{1}H^{N}$ or $^{1}H^{\alpha}$ detection.

The investigation of the major differences between folded and unfolded systems, of how

the unfolded state impacts on nuclear spin relaxation, of which are the critical points that need to be overcome are at the basis of the development of the new experiments described in this thesis. As a conclusion, in collaboration with researches from CNRS and Bruker, a chapter for the book "High resolution tools to understand the functional role of protein intrinsic disorder", edited by Springer, has been written (Paragraph 5.1.4).

NMR was then exploited for the study of IDPs of high biomedical relevance: E7 from HPV 16 (Paragraph 5.2.1 and 5.2.2) and E1A from HAdV (Paragraph 5.2.3). Despite documented attempts, the characterization of these proteins either through NMR or X-ray crystallography failed due to their heterogeneous structural and dynamic properties of the two proteins. Therefore these two proteins provided good targets to test the novel NMR methods optimized for the study of IDPs. The highly flexible regions of both proteins were readily characterized by a combination of ^{1}H and ^{13}C detected experiments, despite challenging features such as the highly repetitive sequences, the high amount of proline residues present in the primary sequence and, particularly for E1A also the quite large size for NMR investigation. The study of these proteins however also revealed new challenges for further developments of NMR methods. Indeed a more structured module, the CR3 domain containing a zinc binding motif, with common elements in the two proteins, initially escaped detection through NMR. A combination of sample optimization, molecular biology and simple and sensitive NMR experiments successfully contributed to the characterization of most (70%) of the CR3 domain of E7 from HPV 16 (Paragraph 5.2.2) and lays the basis for the characterization of the analogous module in E1A.

New challenges are thus emerging for further development of NMR methods to be able to also focus on proteins characterized by significant structural and dynamic heterogeneity hosting both largely disordered as well as partially ordered domains. Finally, the sequence specific assignment and characterization of the E7 from HPV 16 and E1A from HAdV open the way to understanding the molecular basis responsible for their function. The two proteins have the common feature of their overall organization in conserved regions and harbor many common SLiMs responsible for their role as molecular hubs in biological processes.

5.1 NMR methods to study IDPs

5.1.1

NMR spectroscopic studies of intrinsically disordered proteins at near-physiological conditions

Sergio Gil[a], Tomáš Hošek[b], Zsófia Sólyom[c], Rainer Kümmerle[a], Bernhard Brutscher[c], Roberta Pierattelli[b], Isabella C. Felli[b]

[a] *Bruker BioSpin AG, Industriestrasse 26, Fällanden, 8117, Switzerland*

[b] *CERM and Department of Chemistry Ugo Schiff, University of Florence, via Luigi Sacconi 6, Sesto Fiorentino, 50019, Italy*

[c] *Institut de Biologie Structurale, Université Grenoble 1, CNRS, CEA Rue Jules Horowitz 41, Grenoble Cedex 1, 38027, France*

Published in **2013** in *Angewandte Chemie Int. Ed.*, 52(45), 11808-12

NMR Spectroscopy of Proteins

DOI: 10.1002/anie.201304272

NMR Spectroscopic Studies of Intrinsically Disordered Proteins at Near-Physiological Conditions**

*Sergio Gil, Tomáš Hošek, Zsofia Solyom, Rainer Kümmerle, Bernhard Brutscher, Roberta Pierattelli, and Isabella C. Felli**

Intrinsically disordered proteins (IDPs) have recently attracted the attention of the scientific community because of their peculiar features that expand our view of how protein function is determined by the conformational properties of a polypeptide chain. The discovery of numerous physiological functions performed by IDPs has challenged the traditional structure–function paradigm.[1–4] The lack of a unique stable 3D structure and the high extent of local mobility provide functional advantages to IDPs in terms of structural plasticity and binding promiscuity.

Among the experimental techniques available to obtain atomic-resolution information on IDPs, solution-state NMR spectroscopy plays a predominant role. As a spectroscopic method, NMR is equally well applicable to structured as well as highly disordered proteins. However, the absence of a stable 3D structure and the highly dynamic nature of IDPs causes averaging of NMR chemical shifts, resulting in extensive cross-peak overlap in the NMR spectra.[5–9] Furthermore, fast hydrogen exchange of the solvent-exposed amide and other labile protons with water protons causes extensive line broadening or even complete disappearance of the corresponding resonances in the NMR spectra. Therefore, new NMR methods need to be derived or conventional techniques developed for folded proteins need to be adapted to account for the particular spectroscopic properties of IDPs.

To reduce the spectral overlap problem, uniform isotopic enrichment of the IDP with stable isotopes (^{13}C, ^{15}N) is mandatory. Indeed, ^{13}C and ^{15}N nuclei are characterized by an increased chemical shift dispersion with respect to protons. Therefore, proton-detected triple-resonance experiments[10,11] provide the spectral resolution required for site-resolved investigation of IDP structure, dynamics, and interaction modes.[12–14] Recently, the advent of more sensitive NMR instrumentation has brought carbon detection in the suitable sensitivity range for biomolecular NMR applications.[15–17] ^{13}C-Detected exclusively heteronuclear NMR experiments have been developed and offer valuable ways for the characterization of IDPs.[18–21]

The impact of hydrogen protein–solvent exchange processes on the NMR spectra of IDPs has never been discussed in detail, despite its importance, as most NMR experiments used to characterize IDPs are based on amide proton detection. Hydrogen exchange rates are highly dependent on the sample pH and temperature,[22] and changes in these two parameters can thus strongly influence the quality of the spectra. Inspecting the IDP literature reveals that most NMR studies have been performed at low temperatures and (slightly) acidic pH to reduce exchange-induced line broadening in the ^{1}H–^{15}N spectra, and in triple-resonance experiments that are based on amide ^{1}H detection.

The ensemble of conformers that an IDP adopts in solution and its ability to bind to molecular partners, may be strongly dependent on the environmental conditions. Therefore, it is important to have NMR methods to access atomic-resolution information under conditions that more closely approach physiological conditions (neutral pH, body temperature). In this context, ^{13}C detection provides a valuable method of investigation as ^{13}C nuclei are not only characterized by a good chemical shift dispersion even in absence of a stable 3D structure but they are also insensitive to hydrogen-exchange-induced line broadening.[23–27] With increasing temperature and pH, the quality of 2D ^{1}H–^{15}N correlation spectra deteriorates owing to the more efficient hydrogen-exchange mechanism, resulting in increasing number of peaks that are extensively line broadened, while at the same time the quality of the ^{13}C-detected ^{13}CO–^{15}N spectra is maintained, if not improved. This is illustrated in Figure 1, which shows ^{1}H–^{15}N and ^{13}CO–^{15}N spectra recorded in the temperature range of 285.7 K to 304.8 K (pH 7.4) for the IDP α-synuclein, a 140 residue human protein implicated in neurodegenerative diseases.[28] This comparison, which is a general behavior observed for IDPs, demonstrates that ^{13}C-start ^{13}C-detected experiments enable us to recover atomic resolution information that is not available from amide ^{1}H-detected experiments, thus opening new possibilities for the characterization of IDPs under close to physiological sample conditions.

[*] T. Hošek, Prof. R. Pierattelli, Prof. I. C. Felli
CERM and Department of Chemistry "Ugo Schiff"
University of Florence
Via Luigi Sacconi 6, 50019 Sesto Fiorentino, Florence (Italy)
E-mail: felli@cerm.unifi.it
Dr. S. Gil, Dr. R. Kümmerle
Bruker BioSpin AG
Industriestrasse 26, 8117 Fällanden (Switzerland)
Z. Solyom, Dr. B. Brutscher
Institut de Biologie Structurale, Université Grenoble 1, CNRS, CEA
Rue Jules Horowitz 41, 38027 Grenoble Cedex 1 (France)

[**] We thank Drs. Melanie Schwarten and Dieter Willbold (FZ Jülich (Germany)) for their contributions to the NS5A phosphorylation study, Leonardo Gonnelli (CERM) for the α-synuclein samples, and Dr. Klaus-Peter Neidig and Dr. Wolfgang Bermel (Bruker BioSpin GmbH) for stimulating discussions. This work has been supported in part by the EC Project BioNMR (Contract no 261863). S.G., T.H., and Z.S. are fellows of the IDPbyNMR Marie Curie action of the EC (Contract no 264257).

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201304272.

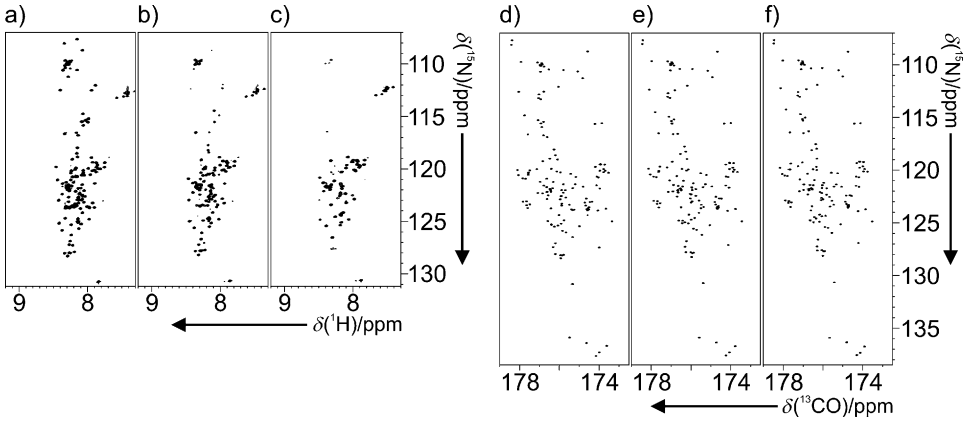


Figure 1. 2D spectra correlating the backbone amide nitrogen either with the directly bound amide proton or with the directly bound carbonyl. a)–c) ^{1}H–^{15}N HSQC and d)–f) ^{13}C–^{15}N CON acquired on α-synuclein at pH 7.4 are shown as a function of increasing temperature: a), d) 285.7 K; b), e) 295.5 K; c), f) 304.8 K. Each spectrum was acquired with one scan per increment, and the same spectral resolution (in Hz) was chosen for the two experiments. Similar behaviour is observed employing different variants of 2D ^{1}H–^{15}N correlation experiments (Supporting Information, Figure S1).

A major drawback of ^{13}C-detected experiments remains their low intrinsic sensitivity owing the approximately four times lower gyromagnetic ratio of ^{13}C with respect to ^{1}H ($\gamma_C \approx \gamma_H/4$). Therefore, it has been proposed to use the larger ^{1}H (instead of ^{13}C) polarization as a starting point of the coherence transfer pathways to increase the sensitivity of ^{13}C-detected experiments.[15,18] The sensitivity of these ^{1}H-start, ^{13}C-detected experiments can be further improved by longitudinal ^{1}H relaxation enhancement techniques.[29–34] These methods are based on the selective manipulation of a subset of proton spins that then relax back to thermodynamic equilibrium much faster owing to an energy transfer mechanism from the excited protons to the protons that remained unperturbed by the pulse sequence. The extent of longitudinal relaxation enhancement obtained by selective manipulation of amide ^{1}H spins in α-synuclein is shown in Figure 2. ^{1}H polarization inversion recovery has been measured for individual amide sites by inverting different pools of ^{1}H spins, followed by a relaxation delay, and a 2D $H^{N\text{-flip}}CON$ readout sequence (Supporting Information, Figure S2) at pH 7.4 and 295.5 K (additional profiles at pH 6.4 are shown in the Supporting Information, Figure S3). Selective inversion of amide protons results in very fast ^{1}H recovery with effective T_1 time constants of about 60 ms (1.4 times the zero-crossing time point), instead of $T_1 \approx 2$ s measured for non-selective ^{1}H inversion. This circa 30-fold decrease in ^{1}H T_1 is impressive, and much higher than what is typically observed for folded proteins with reported selective amide ^{1}H T_1 of 200–400 ms compared to non-selective ^{1}H T_1 of 1.0 to 1.5 s.[35,36] Our inversion recovery data also demonstrate that the major source of ^{1}H relaxation enhancement under these experimental conditions can be attributed to hydrogen-exchange processes rather than ^{1}H–^{1}H dipolar interactions. The latter are responsible for the observed relaxation-enhancement effects in globular proteins,[30–32,34] as well as in IDPs studied at low temperature and acidic pH values.[14]

Inversion recovery curves also show that perturbation of the water resonance dramatically slows down recovery of amide protons to equilibrium. This has a strong impact on the performance of NMR spectra of IDPs, in particular when approaching physiological conditions. Therefore avoiding perturbation of water protons becomes a key general aspect to consider in the design of NMR spectra optimized for the study of IDPs.

The extremely short recovery times observed for amide ^{1}H in α-synuclein at close to physiological conditions prompted us to design new $^{1}H^{N}$-start ^{13}C-detected experiments with minimal perturbation of the water ^{1}H spins that are expected to yield significantly improved sensitivity. A particularly useful example is the 2D $H^{N\text{–BEST}}CON$ experiment shown in Figure 3a that yields "fingerprint" spectra of the protein backbone that can be used for chemical shift mapping purposes. The acronym BEST[31] refers to the use of band-selective amide ^{1}H pulses to achieve longitudinal relaxation enhancement. The BEST technique was preferred over alternative solutions, for example, flip-back approaches, as it performs best with respect to the remaining amount of water ^{1}H polarization after one scan. Because of the long ^{1}H T_1 of water (ca. 3 s at

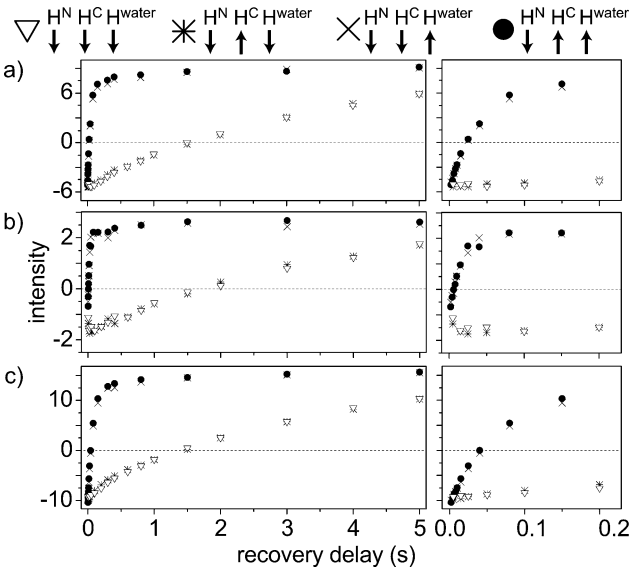


Figure 2. Inversion recovery profiles of amide protons acquired with the variant of the $H^{N\text{–flip}}CON$ (described in the Supporting Information, Figure S2) for selected amino acids of α-synuclein: a) Val 40; b) His 50; c) Leu 100. Different initial conditions are indicated as follows: ▽ non-selective, ⋆ H^{N}–H^{water}-selective, × H^{N}–H^{C}-selective, ● H^{N}-selective.

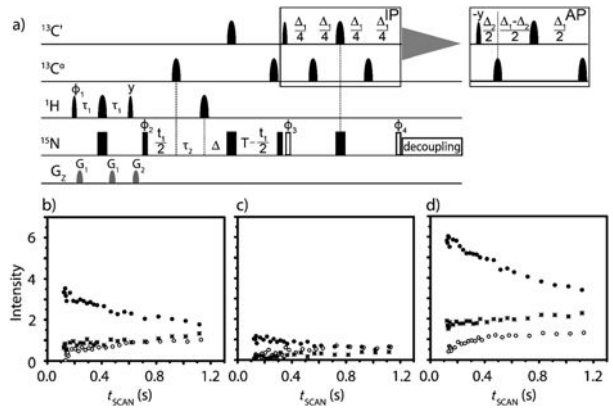


Figure 3. a) $H^{N\text{-}BEST}CON$ pulse sequence. The delays are: $\tau_1 = 1/(4J_{NH}) - 0.5\delta_1 - 0.5\delta_2$, $\tau_2 = 1/(4J_{NH}) - 0.5\delta_2$, $\Delta_1 = 1/(2J_{CON})$, 33.2 ms, $\Delta_2 = 1/(2J_{CO\alpha})$, 9 ms, $\Delta = 1/(4J_{CON}) - 1/(4J_{NH}) - t_1/2$, $T = 1/(4J_{CON})$. The delays δ_1 and δ_2 correspond to the lengths of the PC9 and REBURP pulses that are used for 90° and 180° pulses, respectively. The phase cycle is $\phi_1 = x, -x$; $\phi_2 = 2(x), 2(-x)$; $\phi_3 = 4(x), 4(-x)$; $\phi_4 = 4(y), 4(-y)$; $\phi_{rec} = x, -x, x, -x$. Quadrature detection was obtained by incrementing phase ϕ_2 (t_1) in States-TPPI manner. For the ^{15}N virtual decoupling version, two 90° ^{15}N radio-frequency pulses are applied (white rectangles) instead of the one 90° ^{15}N rf pulse preceding IPAP block and ^{15}N decoupling during acquisition. b)–d) Intensity normalized accordingly to T_{scan} time, for CON (○), $H^{N\text{-}start}CON$ (⋆), and $H^{N\text{-}BEST}CON$ (●) for selected amino acids of α-synuclein, b) Val 40, c) His 50, and d) Leu 100 at pH 7.4 and 295.5 K.

room temperature),[37,38] even minimal perturbation of the water ^1H results in significantly reduced steady-state water ^1H polarization under fast-pulsing conditions, and thus in a reduction of the longitudinal enhancement effect required for optimal sensitivity.

In fact, ^1H recovery starts right after the last ^1H pulse, well before the end of the pulse sequence, in contrast to ^1H-detected experiments that typically end with ^1H pulses. Therefore, at the end of the $H^{N\text{-}BEST}CON$ pulse sequence (including signal acquisition), sufficient ^1H polarization may already have built up to immediately start the next scan without any additional delay. To allow for long ^{13}C acquisition times to take benefit of the favorable transverse relaxation properties of IDPs, and avoid problems related to probe heating owing to ^{15}N decoupling during acquisition, we have implemented virtual ^{15}N decoupling in the pulse sequence.[18,39]

The performance of the new experiment in terms of sensitivity was evaluated by recording a series of 2D ^{13}CO–^{15}N correlation spectra of α-synuclein with different inter-scan delays, ranging from 0 to 1 s, using either $H^{N\text{-}BEST}CON$, the analogous experiment employing non-selective ^1H pulses, $H^{N\text{-}start}CON$, or the ^{13}C-start CON pulse schemes. The results obtained for selected residues of α-synuclein are shown in Figure 3 b–d. The first observation from these data is that starting with ^1H as a polarization source in the $H^{N\text{-}start}CON$ experiment does not result in a significant sensitivity increase with respect to the ^{13}C-start CON version under these experimental conditions. This is mainly explained by the long non-selective ^1H T_1 of about 2 s compared to the ^{13}CO T_1 of about 1 s (estimated value), and by additional ^1H relaxation occurring during the ^1H-start pulse sequence. It is interesting to note how under these conditions, amide ^1H polarization recovery for IDPs may become slower than ^{13}C recovery of carbonyl nuclei. Much larger signal gains are obtained for the $H^{N\text{-}BEST}CON$ experiment that benefits from the extremely short selective recovery times of amide ^1H that are used as starting polarization. The observed sensitivity gains ($H^{N\text{-}BEST}CON$ over ^{13}C-start CON), determined for short recycle times of 150 ms, vary between different sites from a factor of 4 to a factor of 12 (Supporting Information, Figure S4). The example curves shown in Figure 3 b–d have been chosen to be representative of residues characterized by solvent exchange rates of different magnitude. For residues characterized by (relatively) slow exchange under these conditions (for example, Leu 100), a higher sensitivity increase is observed than for residues with faster exchanging amide protons (for example, His 50). This demonstrates that starting from amide ^1H introduces again a dependency on the solvent exchange rates, and that there is an optimal exchange regime for the use of such BEST-type ^1H-start ^{13}C-detected experiments. Maximal sensitivity is achieved if the exchange rate is large with respect to the longitudinal relaxation rate R_1, but still small compared to the transverse relaxation rate R_2 ($1\ s^{-1} < k_{ex} < 10^2\ s^{-1}$). In the presence of even faster hydrogen exchange, experimental variants exploiting ^{13}C,[23,25] or non-exchangeable ^1H,[18] as the starting polarization source are expected to yield higher sensitivity.

The $H^{N\text{-}BEST}CON$ experiment is particularly useful to study the behavior of IDPs inside living cells (in-cell NMR) or to monitor post-translational modifications occurring either in vitro or in vivo, because to be biologically relevant, all of these studies are preferably performed under sample conditions that are close to physiological pH and temperature.

In-cell NMR spectroscopy provides a unique spectroscopic method to investigate the fate of a protein in a cellular context. In particular, it allows changes induced by different cellular stimuli to be monitored or protein spectra obtained for different cell types to be compared.[40–44] A number of techniques, for example bacterial over-expression, injection, electroporation, or direct expression of proteins in human cells, have been developed during recent years that allow to introduce isotopic labeled protein into various cell types.[40–44] The extensive line broadening that is typical for in-cell spectra, in combination with the drastically reduced chemical shift dispersion and high solvent exchange rates of IDPs, has a strong negative impact on the amount of residue-specific information that can be obtained through 2D ^1H–^{15}N correlation experiments. ^{13}C-detected experiments offer a valuable solution to this problem thanks to the higher chemical shift dispersion and reduced sensitivity to exchange broadening.[25,45] However, high sensitivity is required to ensure short experimental times to cope with the limited lifetime of meaningful in-cell NMR samples. The use of ^1H as a starting polarization source as well as longitudinal relaxation enhancement are important features for the study of IDPs in cells through ^{13}C-detected exclusively heteronuclear NMR experiments.[46] An example of an in-cell $H^{N\text{-}BEST}CON$ spectrum of α-synuclein over-expressed in *E. coli* cells is shown in Figure 4. The inset clearly shows how the $H^{N\text{-}BEST}CON$ experiment enables to detect correlations that were lost

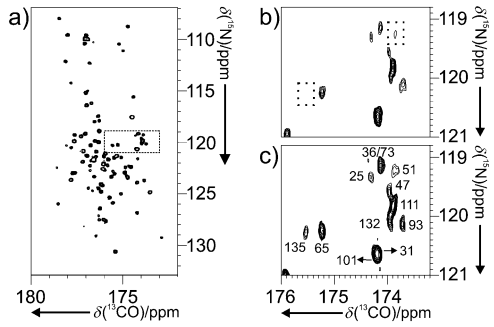


Figure 4. a) 2D ^{13}C–^{15}N $H^{N\text{-}BEST}CON$ spectrum of α-synuclein in *E. coli* cells acquired in 20 min. The region of the spectrum enclosed by a rectangular box is enlarged to highlight how c) $H^{N-BEST}CON$ enables a better sensitivity to be obtained compared to b) $H^{N-flip}CON$ and to detect correlations that otherwise would not have been detected in the same experimental time.

under the spectral noise in the corresponding H^{N-flip} implementation of the experiment.

Another interesting application of NMR spectroscopy at nearly physiological conditions consists in following post-translational modifications modulating protein function. Among them, recent studies have focused on monitoring phosphorylation events involving intrinsically disordered proteins.[47–49] Also in this case, ^{13}C-detected NMR experiments complement well the ^{1}H-detected experiments for obtaining site-resolved information on phosoporylation events in intrinsically disordered proteins (for an example, see the Supporting Information, Figure S5).

In conclusion, we have shown here that ^{13}C-start ^{13}C-detected experiments do not suffer from fast hydrogen exchange between amide and solvent protons in IDP samples studied at close to physiological conditions, thus enabling us to recover information that would be difficult or even impossible to obtain through amide ^{1}H-detected experiments. Furthermore, in favourable cases the fast hydrogen exchange rates can even be turned into a spectroscopic advantage. By combining longitudinal ^{1}H relaxation optimized BEST-type techniques with ^{13}C-direct detection pulse schemes, important sensitivity improvements can be achieved, and experimental times can be significantly reduced. This opens up new applications for monitoring chemical shift changes in IDPs upon interaction to a binding partner, chemical modification, or by changing the environment, under sample conditions that were inaccessible by conventional techniques. This concept can be easily extended to other $^{1}H^{N}$-start ^{13}C-detected experiments. Furthermore, ^{13}C-detected NMR experiments can be used to obtain semi-quantitative information about hydrogen-exchange rates without directly detecting the exchangeable amide ^{1}H by quantifying the signal enhancement observed in $H^{N-BEST}CON$ with respect to $H^{N\text{-}start}CON$ spectra.[18,33] This may become useful to monitor changes in the exchange rates after for example, some binding event, reflecting a change in the solvent accessibility at the binding site of the IDP.

Received: May 17, 2013
Revised: July 30, 2013
Published online: September 20, 2013

Keywords: intrinsically disordered proteins · longitudinal relaxation enhancement · NMR spectroscopy · α-synuclein

[1] P. E. Wright, H. J. Dyson, *J. Mol. Biol.* **1999**, *293*, 321–331.
[2] P. Tompa, M. Fuxreiter, *Trends Biochem. Sci.* **2008**, *33*, 2–8.
[3] P. Tompa, *Structure and function of intrinsically disordered proteins*, Taylor and Francis Group, Boca Raton, FL, **2009**.
[4] V. N. Uversky, A. K. Dunker, *Biochim. Biophys. Acta Proteins Proteomics* **2010**, *1804*, 1231–1264.
[5] D. Neri, G. Wider, K. Wüthrich, *Proc. Natl. Acad. Sci. USA* **1989**, *89*, 4397–4401.
[6] H. J. Dyson, P. E. Wright, *Methods Enzymol.* **2001**, *339*, 258–271.
[7] H. Zhang, S. Neal, D. S. Wishart, *J. Biomol. NMR* **2003**, *25*, 173–195.
[8] D. Eliezer, *Curr. Opin. Struct. Biol.* **2009**, *19*, 23–30.
[9] J. Marsh, S. A. Teichmann, J. D. Forman-Kay, *Curr. Opin. Struct. Biol.* **2012**, *22*, 643–650.
[10] M. Sattler, J. Schleucher, C. Griesinger, *Prog. Nucl. Magn. Reson. Spectrosc.* **1999**, *34*, 93–158.
[11] A. Bax, S. Grzesiek, *Acc. Chem. Res.* **1993**, *26*, 131–138.
[12] M. D. Mukrasch, S. Bibow, J. Korukottu, S. Jeganathan, J. Biernat, C. Griesinger, E. Mendelkow, M. Zweckstetter, *PLoS Biol.* **2009**, *7*, e34.
[13] V. Motáčková, J. Nováček, A. Zawadzka-Kazimierczuk, K. Kazimierczuk, L. Žídek, H. Šanderová, L. Krásný, W. Koźmiński, V. Sklenář, *J. Biomol. NMR* **2010**, *48*, 169–177.
[14] Z. Solyom, M. Schwarten, L. Geist, R. Konrat, D. Willbold, B. Brutscher, *J. Biomol. NMR* **2013**, *55*, 311–321.
[15] Z. Serber, C. Richter, D. Moskau, J.-M. Boehlen, T. Gerfin, D. Marek, M. Haeberli, L. Baselgia, F. Laukien, A. S. Stern, J. C. Hoch, V. Dötsch, *J. Am. Chem. Soc.* **2000**, *122*, 3554–3555.
[16] I. Bertini, L. Duma, I. C. Felli, M. Fey, C. Luchinat, R. Pierattelli, P. R. Vasos, *Angew. Chem.* **2004**, *116*, 2307–2309; *Angew. Chem. Int. Ed.* **2004**, *43*, 2257–2259.
[17] H. Kovacs, D. Moskau, M. Spraul, *Prog. Nucl. Magn. Reson. Spectrosc.* **2005**, *46*, 131–155.
[18] W. Bermel, I. Bertini, V. Csizmok, I. C. Felli, R. Pierattelli, P. Tompa, *J. Magn. Reson.* **2009**, *198*, 275–281.
[19] W. Bermel, I. Bertini, I. C. Felli, R. Peruzzini, R. Pierattelli, *ChemPhysChem* **2010**, *11*, 689–695.
[20] J. Nováček, A. Zawadzka-Kazimierczuk, V. Papoušková, L. Žídek, H. Šanderová, L. Krásny, W. Koźmiński, V. Sklenář, *J. Biomol. NMR* **2011**, *50*, 1–11.
[21] W. Bermel, I. Bertini, L. Gonnelli, I. C. Felli, W. Kozminski, A. Piai, R. Pierattelli, J. Stanek, *J. Biomol. NMR* **2012**, *53*, 293–301.
[22] Y. W. Bai, J. S. Milne, L. Mayne, S. W. Englander, *Proteins Struct. Funct. Genet.* **1993**, *17*, 75–86.
[23] W. Bermel, I. Bertini, I. C. Felli, R. Kümmerle, R. Pierattelli, *J. Magn. Reson.* **2006**, *178*, 56–64.
[24] I. Bertini, I. C. Felli, L. Gonnelli, R. Pierattelli, Z. Spyranti, G. A. Spyroulias, *J. Biomol. NMR* **2006**, *36*, 111–122.
[25] W. Bermel, I. Bertini, I. C. Felli, Y.-M. Lee, C. Luchinat, R. Pierattelli, *J. Am. Chem. Soc.* **2006**, *128*, 3918–3919.
[26] S. T. Hsu, C. W. Bertoncini, C. M. Dobson, *J. Am. Chem. Soc.* **2009**, *131*, 7222–7223.
[27] L. Skora, S. Becker, M. Zweckstetter, *J. Am. Chem. Soc.* **2010**, *132*, 9223–9225.
[28] P. H. Weinreb, W. G. Zhen, A. W. Poon, K. A. Conway, P. T. Lansbury, Jr., *Biochemistry* **1996**, *35*, 13709–13715.
[29] K. Pervushin, B. Vogeli, A. Eletsky, *J. Am. Chem. Soc.* **2002**, *124*, 12898–12902.
[30] P. Schanda, B. Brutscher, *J. Am. Chem. Soc.* **2005**, *127*, 8014–8015.

[31] P. Schanda, H. Van Melckebeke, B. Brutscher, *J. Am. Chem. Soc.* **2006**, *128*, 9042 – 9043.
[32] E. Lescop, P. Schanda, B. Brutscher, *J. Magn. Reson.* **2007**, *187*, 163 – 169.
[33] W. Bermel, I. Bertini, I. C. Felli, R. Pierattelli, *J. Am. Chem. Soc.* **2009**, *131*, 15339 – 15345.
[34] A. Favier, B. Brutscher, *J. Biomol. NMR* **2011**, *49*, 9 – 15.
[35] B. Brutscher, P. Schanda in *Encyclopedia of NMR* (Eds.: M. Grant, R. K. Harris), J. Wiley and Sons, Chichester, **2009**.
[36] P. Schanda, *Prog. Nucl. Magn. Reson. Spectrosc.* **2009**, *55*, 238 – 265.
[37] H. Y. Carr, E. M. Purcell, *Phys. Rev.* **1954**, *94*, 630 – 638.
[38] H. J. Simpson, H. Y. Carr, *Phys. Rev.* **1958**, *111*, 1201 – 1202.
[39] T. Kern, P. Schanda, B. Brutscher, *J. Magn. Reson.* **2008**, *190*, 333 – 338.
[40] S. Reckel, F. Lohr, V. Dotsch, *ChemBioChem* **2005**, *6*, 1601 – 1606.
[41] K. Inomata, A. Ohno, H. Tochio, S. Isogai, T. Tenno, I. Nakase, T. Takeuchi, S. Futaki, Y. Ito, H. Hirokai, M. Shirakawa, *Nature* **2009**, *458*, 106 – 109.
[42] S. Liokatis, A. Stuetzer, S. J. Elsaesser, F. X. Theillet, R. Klingberg, B. van Rossum, D. Schwarzer, C. D. Allis, W. Fischle, P. Selenko, *Nat. Struct. Mol. Biol.* **2012**, *19*, 819 – 823.
[43] F. X. Theillet, C. Smet-Nocca, S. Liokatis, R. Thongwichian, J. Kosten, M. K. Yoon, R. W. Kriwacki, I. Landrieu, G. Lippens, P. Selenko, *J. Biomol. NMR* **2012**, *54*, 217 – 236.
[44] L. Banci, L. Barbieri, I. Bertini, E. Luchinat, E. Secci, Y. Zhao, A. R. Aricescu, *Nat. Chem. Biol.* **2013**, *9*, 297 – 299.
[45] A. Binolfi, F. X. Theillet, P. Selenko, *Biochem. Soc. Trans.* **2012**, *40*, 950 – 954.
[46] I. Bertini, I. C. Felli, L. Gonnelli, V. M. V. Kumar, R. Pierattelli, *Angew. Chem.* **2011**, *123*, 2387 – 2389; *Angew. Chem. Int. Ed.* **2011**, *50*, 2339 – 2341.
[47] G. Lippens, L. Amniai, M. R. Wieckowski, A. Sillen, A. Leroy, I. Landrieu, *Biochem. Soc. Trans.* **2012**, *40*, 698 – 703.
[48] F. X. Theillet, H. M. Rose, S. Liokatis, A. Binolfi, R. Thongwichian, M. Stuiver, P. Selenko, *Nat. Protoc.* **2013**, *8*, 1416 – 1432.
[49] I. Amata, M. Maffei, A. Igea, M. Gay, M. Vilaseca, A. R. Nebreda, M. Pons, *ChemBioChem* **2013**, DOI: 10.1002/cbic.201300139.

Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

69451 Weinheim, Germany

NMR Spectroscopic Studies of Intrinsically Disordered Proteins at Near-Physiological Conditions**

*Sergio Gil, Tomáš Hošek, Zsofia Solyom, Rainer Kümmerle, Bernhard Brutscher, Roberta Pierattelli, and Isabella C. Felli**

anie_201304272_sm_miscellaneous_information.pdf

Protein samples

α–synuclein was expressed and purified as previously described.[1] NMR experiments were acquired on a 1.0 mM ^{13}C, ^{15}N- labeled sample (550 µl) in 20 mM Pi buffer and 200 mM NaCl at pH 7.4; 5% D_2O was added for the lock. To assess the effect of pH on the inversion recovery profiles, a sample in exactly the same conditions except for pH, that was set at 6.4, was also prepared.

In-cell NMR experiments were acquired on *E.coli* cells overexpressing α–synuclein, grown in ^{13}C, ^{15}N isotopically enriched minimal medium after induction of protein expression. For the preparation of the NMR sample, a single colony from a freshly transformed BL21(DE3) gold *E. coli* harbouring the expression plasmid was inoculated in 70 ml of LB medium supplemented with 100 µg/ml of carbenicillin in a 250 ml flask and grown overnight in thermostatised shaker at 27°C at 380 rpm (20 mm oscillation amplitude). The day after the cell density reached OD_{600} between 4 and 6. 35 ml of the culture (OD_{600} =6) were centrifuged gently (2000 x g) at 4°C for 20 minutes. The cells were re-suspended in 50 ml of labeled medium in a 250 ml flask and the flask was put back in the shaker (37°C 380 RPM). After 10-15 minutes of temperature recovery the culture was induced by addition of IPTG to a final concentration 1 mM. Cells were collected 4 hours after the induction. Two NMR samples were prepared from the same flask. Cells were centrifuged (2000 x g) at 4°C for 20 minutes and then re-suspended using 300 -500 µl of their culture medium so that the amount of cells was 50% of the entire sample volume (600 µl). A 10% v/v of deuterium oxide was added for lock circuit. The cells viability was checked by spreading the same amount of sample on LB agar plate containing 50 µg/l carbenicillin before and after the NMR experiments. The absence of protein leakage to the culture medium was verified. The cells were removed from the NMR tube and centrifuged at 1500 x g for 10 minutes. The supernatant was further centrifuged at 10000 x g for additional 10 min and checked by NMR analysis.

NS5A was expressed and purified as previously described.[2] NMR experiments were acquired on a 100 µM NS5A sample (300 µl) in a buffer solution containing 50 mM Tris-HCl, 10 mM $MgCl_2$, 5 mM DTT and 5 mM ATP. The CK1 kinase (8 µl) was added to follow the phosphorylation process (8000 units); 5% D_2O was added for the lock. In order to ensure appropriate conditions for the kinase activity, the reaction has been conducted at 25°C and pH 7.0.

NMR experiments

NMR experiments were performed at 16.4 T on a Bruker Avance spectrometer operating at 700.06 MHz ^1H, 176.03 MHz ^{13}C and 70.9 MHz ^{15}N frequencies, equipped with a triple-resonance cryogenically cooled probehead optimized for ^{13}C-direct detection.

The new pulse schemes, the inversion recovery $H^{N\text{-}flip}CON$ and the $H^{N\text{-}BEST}CON$ are illustrated in Figure S2 and in Figure 3 in the main text respectively. All pulses were applied with phase x, unless otherwise specified. In all the panels reported, unless otherwise specified, band-selective ^{13}C pulses are denoted by rounded shapes and rectangular symbols represent hard pulses. The wide and narrow pulses correspond to π and π/2 flip angles. Pulsed field gradients (PFG line) are also indicated by rounded shapes. The common parameters used for all experiments are described in detail in this paragraph, while those specific for each experiment are reported in the captions of the figures.

Unless otherwise specified, for ^{13}C band-selective π/2 and π pulses, Q5 (or time reversed Q5) and Q3 shapes[3] were used with durations of 300 and 220 µs, respectively. The ^1H carrier was placed at 4.7 ppm and the ^{15}N carrier was placed at 118 ppm or at 122.5 ppm to also include signals of prolines. Adiabatic π pulse was used to invert both CO and C^{ali} [4] (500 µs, 25% smoothing, 80 kHz sweep, 11.3 kHz strength). The change in the position of the ^{13}C carrier (39 ppm for C^{ali}, 55 ppm for C^{α} and 173 ppm for CO, 100 ppm for the adiabatic pulse) is indicated by vertical arrows. All proton-selective pulses were applied centered at 9 ppm, covering a bandwidth of 4 ppm with PC9 shaped pulses for π/2 [5] and Reburp shaped

pulses for π.[6] Decoupling of ^{1}H and ^{15}N (when appropriate) was achieved with Waltz-16 [7] (1.7 kHz) and Garp-4 [8] (1.0 kHz) sequences, respectively. All the experiments employ the IPAP approach to suppress the C^{α}-CO coupling in the direct acquisition dimension[9-12] where the in-phase (IP) and antiphase (AP) components are acquired and stored separately, doubling the number of FIDs recorded in one of the indirect dimensions.[13] To allow for long ^{13}C acquisition time, ^{15}N virtual decoupling was implemented as previously described.[14] The phase cycle, the method used for quadrature detection, the duration of the delays shown in the pulse sequences, and the relative strengths of the gradients used (all gradients had a duration of 1ms and sine shape) are reported case-by-case.

The 2D spectra correlating the backbone amide nitrogen either with the directly bound amide proton or with the directly bound carbonyl shown in Figure 1 of the main text (^{1}H-^{15}N HSQC [15;16]and ^{13}C-^{15}N CON[17]) were acquired on α-synuclein at pH 7.4 as a function of increasing temperature (285.7K, 295.5K and 304.8K). The ^{1}H-^{15}N HSQC experiment used in Figure 1 was the sensitivity improvement echo/antiecho version which also includes a water flip-back pulse [15;16]. Each spectrum was acquired with one scan per increment with enough increments to obtain the same resolution in the indirect dimension (512 increments for a spectral width of 26 ppm in the 2D ^{1}H-^{15}N HSQC and 688 increments for a spectral width of 35 ppm in the 2D ^{13}C-^{15}N CON; in the latter case each increment needs to be repeated twice to perform homonuclear virtual decoupling). The interscan delay was set to 1 s and 1.2 s respectively for a total duration of 12 min and 10 s, and 35 min and 18 s for the ^{1}H-^{15}N HSQC and ^{13}C-^{15}N CON, respectively.
Repeating the acquisition of 2D ^{1}H-^{15}N correlation experiments with two scans per increment to compare spectra acquired in the same amount of time, as well as with different experimental variants [18;19] the overall result does not change: with increasing temperature the quality of the 2D ^{1}H-^{15}N spectrum decreases while the quality of the 2D ^{13}C-^{15}N spectrum remains unchanged or even improves and all the cross peaks can be detected.

The reference pulse sequence $H^{N\text{-flip}}$CON, preceded by different variants for inversion recovery (inversion-recovery $H^{N\text{-flip}}$CON), is shown in Figure S2. Spectra were acquired at 295.5 K with standard parameters, using an interscan delay of 5 s to ensure complete recovery to thermal equilibrium, acquisition time of 100 ms with four scans per increment. In the ^{15}N dimension, 200 increments were acquired for a spectral width of 35 ppm resulting in a maximal evolution time of 40.3 ms. Spectra were recorded with 16 relaxation delays, ranging from 0.05 to 5 s in non-selective and H^{N}-H^{water}-selective and 0.02 to 5 s in H^{N}-selective and H^{N}-H^{ali}-selective. The same inversion recovery profiles were acquired also at pH 6.4 and at 285.7 K, all other experimental conditions being the similar (same sample conditions, the experiments were acquired with 128 increments in the ^{15}N dimension for a spectral width 35 ppm resulting in a maximal evolution time of 25.7 ms, and for 13 different relaxation delays ranging from 0.05 to 5 s in non-selective and H^{N}-H^{water}-selective and 0.02 to 5 s in H^{N}-selective and H^{N}-H^{ali}-selective). All relaxation curves were fitted to a mono-exponential function using the Bruker Dynamics Center 2.1.1 program.

Sensitivity curves of $H^{N\text{-BEST}}$CON, $H^{N\text{-start}}$CON and ^{13}C-start CON were acquired at 295.5K with 24 recycle delays from 0.005 to 1s for a total duration of 3 hours and half for the series. Spectra were acquired with an acquisition time of 50 ms, with two scans per increment. In the ^{15}N dimension, 128 increments were acquired for a spectral width of 35 ppm resulting in a maximal evolution time of 25.8 ms. All proton-selective pulses were applied centered at 9 ppm, covering a bandwidth of 4 ppm with PC9 shaped pulses for π/2[5] and Reburp shaped pulses for π rotations.[6]

The in-cell $H^{N\text{-BEST}}$CON spectrum acquired on *E.coli* cells overexpressing α-synuclein was acquired with 16 scans per increment and 116 increments for a spectral width of 26 ppm and maximal evolution time 31.4 ms. The relaxation delay was 160 ms, and the acquisition time 66.4 ms. ^{15}N decoupling was applied during acquisition. The total duration of the experiment was 23 minutes. For comparison the $H^{N\text{-flip}}$CON was acquired with the same parameters.

The series of $H^{N\text{-}BEST}CON$ and CON spectra on NS5A to follow the phosphorylation reaction were acquired at 298.0 K. $H^{N\text{-}BEST}CON$ with 200 scans per increment and 110 increments for a spectral width of 24 ppm, maximal evolution time 32.3 ms and recycle delay 50 ms while CON were acquired with 60 scans per increment and 128 increments for a spectral width of 34 ppm, maximal evolution time of 26.5 ms and recycle delay 1s. In the $H^{N\text{-}BEST}CON$ experiment each increment needs to be repeated four times to perform simultaneous ^{13}C homonuclear and ^{15}N virtual decoupling while for the CON experiment they need to be repeated twice to perform ^{13}C homonuclear virtual decoupling. The total duration of each CON experiment was 6 h 11 min and 13 s while each $H^{N\text{-}BEST}CON$ spectrum was acquired for a total duration of 6 h 4 min and 55 s. After addition of the kinase (start of the reaction) the CON and $H^{N\text{-}BEST}CON$ spectra were acquired alternatingly, seven times each, as long as changes (due to phosphorylation) could be observed in the acquired spectra.

Experimental data were acquired using Bruker TopSpin 1.3 software and processed with Bruker TopSpin 3.1 software.

Figure S1. The 2D spectra correlating the backbone amide nitrogen either with the directly bound amide proton or with the directly bound carbonyl ^{1}H-^{15}N HSQC [15;16] and ^{13}C-^{15}N CON[17]), acquired on α-synuclein at pH 7.4, are shown as a function of increasing temperature (from left to right 285.7K, 295.5K and 304.8K). The spectra were acquired in the same conditions as reported for Figure 1, the only difference being that the sensitivity improved 2D ^{1}H-^{15}N HSQC was acquired with two scans per increment in order to compare ^{1}H-detected and ^{13}C-detected data obtained in comparable amount of time (top panels on the left-hand side). In addition, two different variants of the 2D ^{1}H-^{15}N HSQC experiments, namely a different variant of the 2D water flip-back ^{1}H-^{15}N HSQC experiment (FHSQC, middle panels on the left-hand side) and the 2D ^{1}H-^{15}N SOFAST-HMQC experiment (SF-HMQC, bottom panels on the left-hand side) were acquired.[18;19] These additional spectra show that with increasing temperature the quality of the 2D ^{1}H-^{15}N correlation experiments decreases, while that of the 2D ^{13}C-^{15}N CON does not.

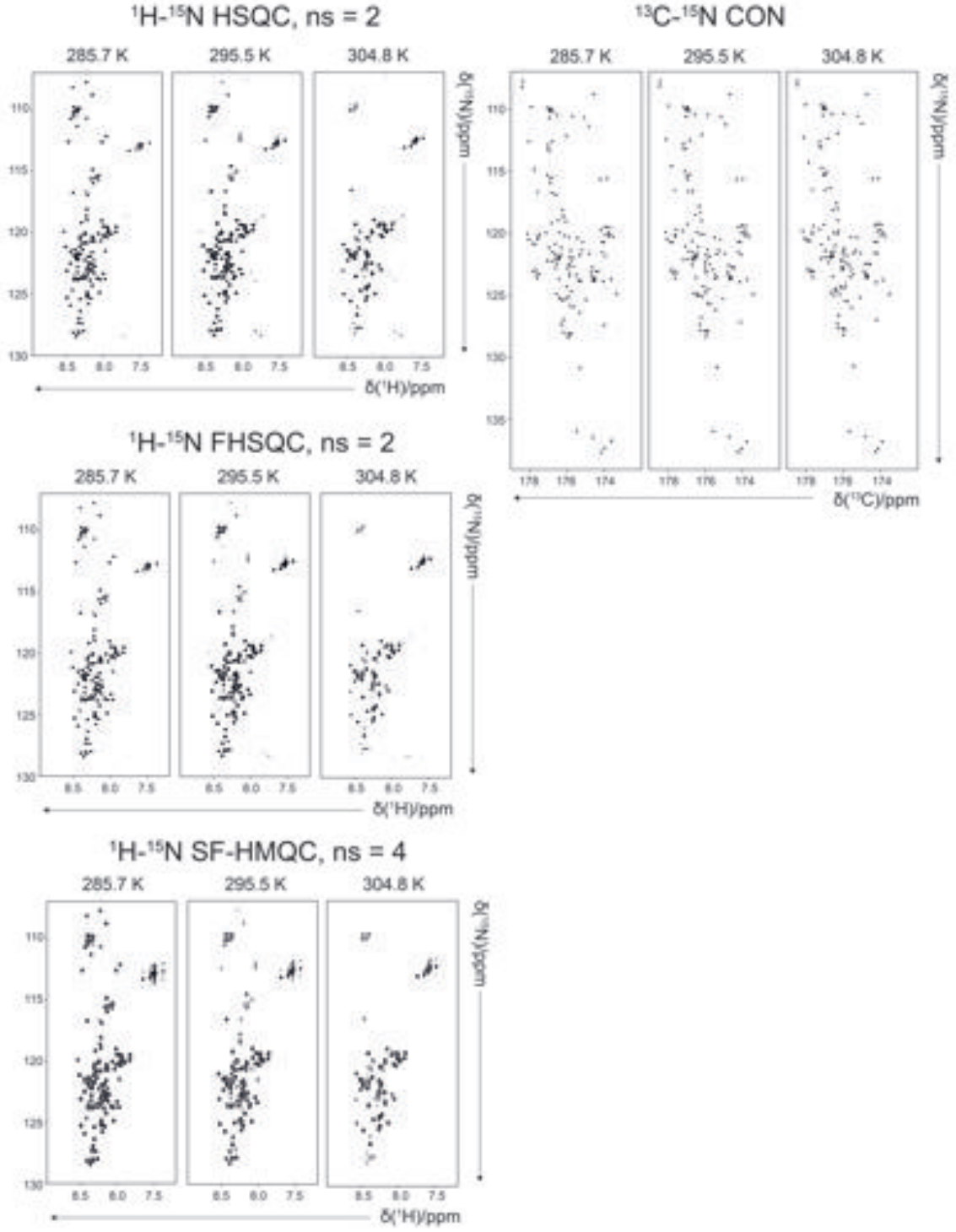


Figure S2. A) Inversion recovery blocks implemented previous to the $H^{N\text{-}flip}CON$ pulse sequence. (a) non-selective (b) H^N-selective, (c) H^N-H^{water}-selective and (d) H^N-H^{ali}-selective inversion recovery blocks. The set of ^1H spins inverted by every block is shown with red arrows. $\Delta_{HN} = 1/(2J_{NH})$, $\Delta_C = 1/(2J_{CH})$, $\Delta' = \Delta_N - \Delta_C$. Gradient power relative to 53.5 G/cm was G1:G2:$G_{strength}$ = 11: 7:1%. A weak gradient was applied during the inversion recovery delay to suppress radiation damping.

B) The $H^{N\text{-}flip}CON$ pulse sequence. The delays are: $\Delta_{NH}=1/(2J_{NH})$, 5.4 ms, $\Delta_1 = 1/(2J_{CON})$, 33.2 ms, $\Delta_2 = 1/(2J_{CO\alpha})$, 9 ms. The phase cycle is ϕ_1= x, -x; ϕ_2= -x, x; ϕ_3 = y,y,-y, -y; ϕ_{IPAP}(IP) = x; ϕ_{IPAP}(AP) = -y; ϕ_{rec} = x, -x,x, -x. Quadrature detection was obtained by incrementing phase ϕ_3 (t_1) in States-TPPI manner. Gradient power relative to 53.5 G/cm was G1:G2:G3:G4 = 5:17:33:50%.

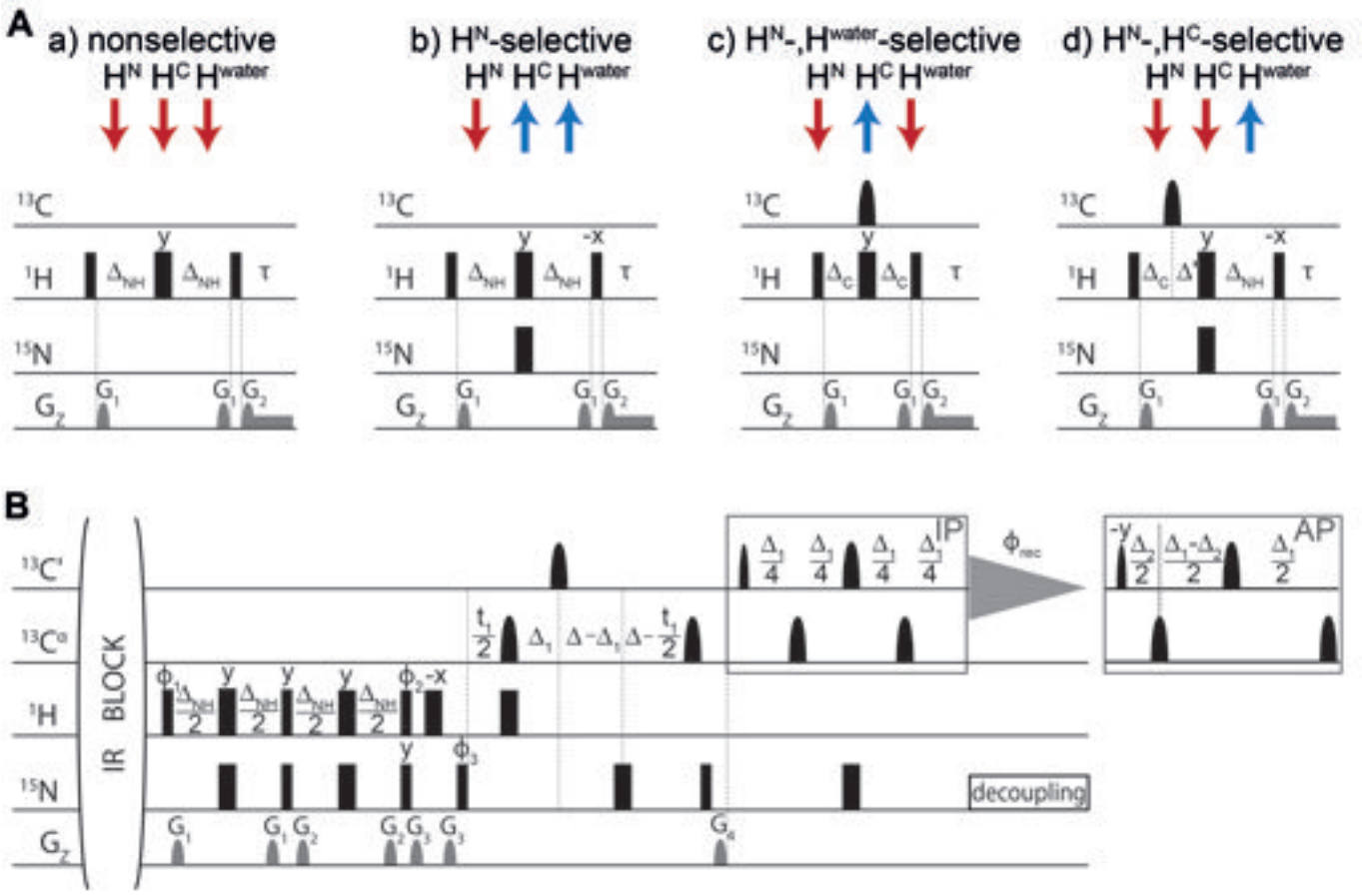


Figure S3. a-c) Inversion recovery profiles of amide protons acquired with the variant of the $H^{N\text{-}flip}CON$ described in Figure S2 for selected amino acids of α-synuclein, (a) Glu 13, (b) Thr 54, (c) Asp 135, at pH 7.4 (left column) and pH 6.4 (middle column), at 295.5 K, and at pH 6.4 at 285.7 K (right column). Different initial conditions are indicated with the following symbols: (♦) Non-selective, (▲) H^N-H^{water}-selective, (■) H^N-H^C-selective, and (●) H^N-selective.
d) The inversion recovery profiles acquired with the pulse sequence described in Figure S2 were fitted to a mono-exponential function to have an estimate of the "effective" T_1 for the amide protons of α-synclein at pH 7.4 and 295.5 K. The distribution of values obtained is shown in form of histograms.

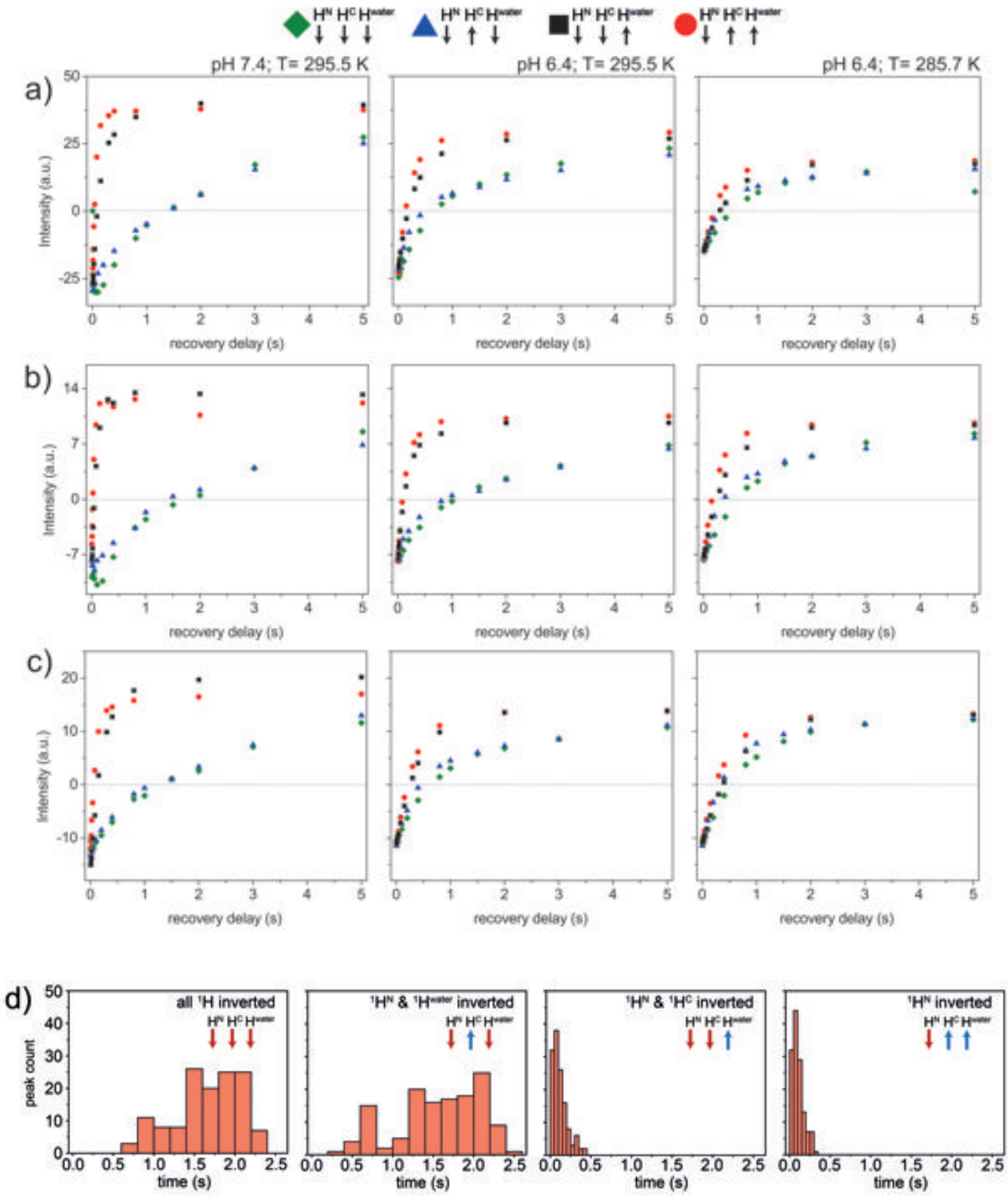


Figure S4. Histogram of the improvement in sensitivity obtained through the $H^{N\text{-}BEST}CON$ experiment respect to the CON experiment for recycle times of 150 ms and of 1 s respectively. Data were acquired on α-synuclein at pH 7.4 and 295.5 K.

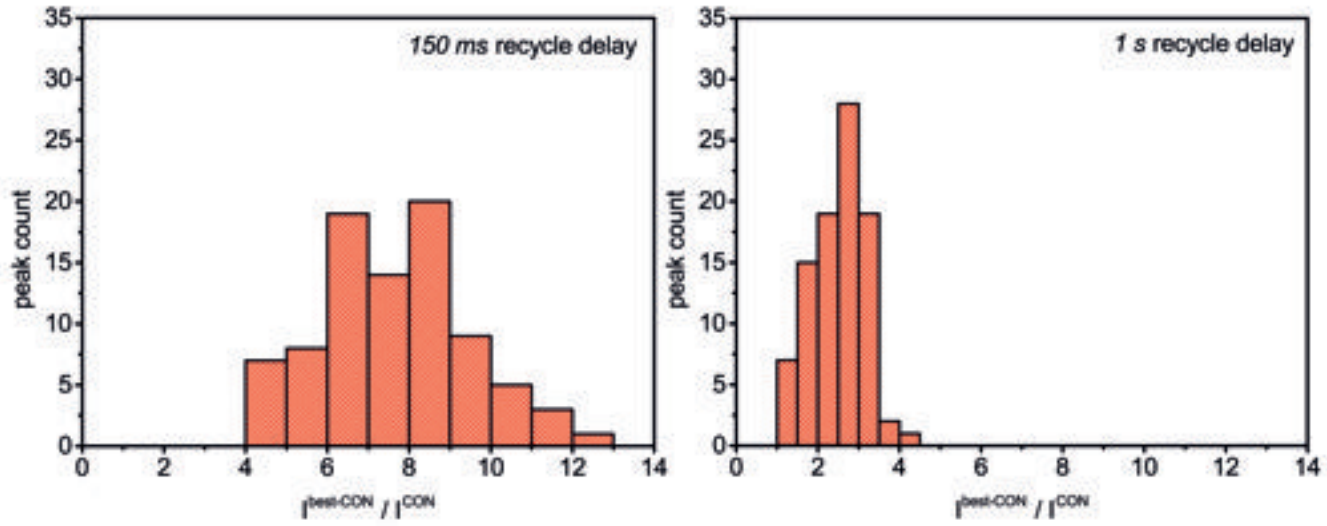
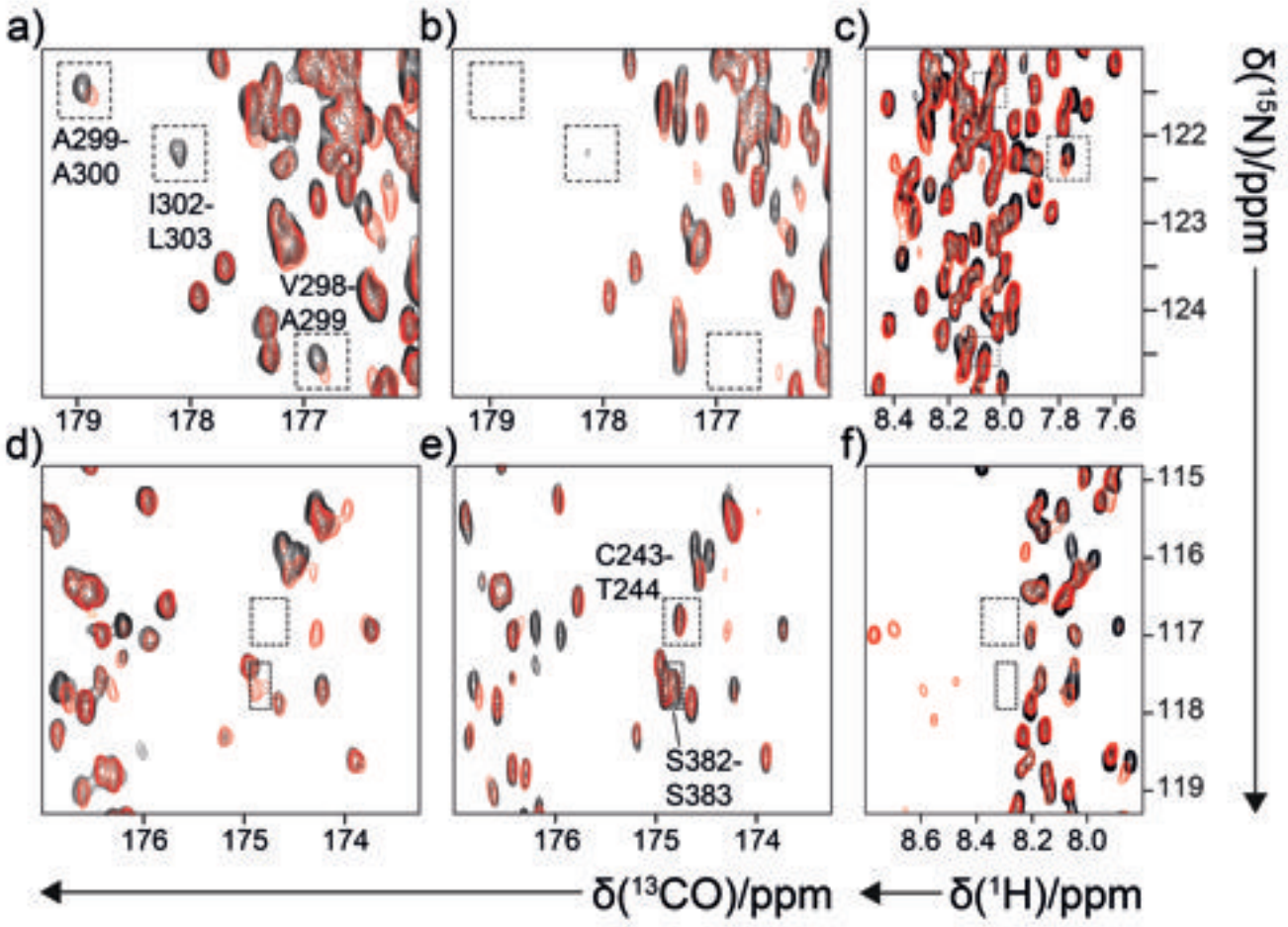


Figure S5. Phosphorylation reaction of NS5A protein from hepatitis C virus by the CK1 kinase monitored by NMR. We have recorded 2D $H^{N\text{-}BEST}CON$, CON, and ^{1}H-^{15}N correlation spectra during the *in vitro* phosphorylation reaction of a NS5A construct comprising residues 191 to 447 that is phosphorylated on multiple serine residues in HCV infected cells[20;21]. In order to ensure appropriate conditions for the kinase activity, the reaction has been conducted at 25°C and pH=7.0. Panels (a) and (d) show spectral regions extracted from 2D $H^{N\text{-}BEST}CON$ spectra recorded before (black) and 4 days after initiating the phosphorylation process by adding CK1 kinase (red). In panels (b) and (e) the corresponding spectral regions extracted from CON experiments recorded before (black) and after (red) phosphorylation are shown. Finally, panels (c) and (f) show spectral regions (same ^{15}N chemical shift range) extracted from conventional sensitivity-enhanced water-flip-back HSQC spectra. All spectra were recorded in an interleaved manner. The ^{1}H-detected experiments still show higher sensitivity than the ^{13}C-detected versions for most of the correlation peaks, but suffer from higher spectral overlap. For the phosphoserines new correlation peaks appear in the spectra, while peak shifts are observed for neigbouring residues and for residues that experience phosphorylation-induced structural changes. The ^{13}CO chemical shift differences thus provide a direct information on changes in secondary structural propensities upon phosphorylation. The spectra illustrate the overall significantly higher signal-to-noise ratio obtained by $H^{N\text{-}BEST}CON$ with respect to CON. This allowed, for example to detect chemical shift changes in the peptide region 298-300 that are due to some conformational changes upon phosphorylation of remote serine residues. The corresponding cross peaks are either very weak or undetectable in the CON spectra. On the other hand these spectra also illustrate the limitations of the $H^{N\text{-}BEST}CON$ approach. Residues with amide protons that exchange very fast with the solvent are not observable in $H^{N\text{-}BEST}CON$ (or ^{1}H-^{15}N HSQC), while they still give rise to correlation peaks in the CON spectra. Overall, the combination of the three techniques allows monitoring the phosphorylation event for the majority of residues within this NS5A construct.



Reference List

[1.] C. Huang, G. Ren, H. Zhou, C. Wang, *Protein.Expr.Purif.* **2005,** *42* 173-177.

[2.] S. Feuerstein, Z. Solyom, A. Aladag, S. Hoffmann, D. Willbold, B. Brutscher, *Biomol.NMR Assign.* **2011,** *5* 241-243.

[3.] L. Emsley, G. Bodenhausen, *Chem.Phys.Lett.* **1990,** *165* 469-476.

[4.] J.-M. Böhlen, G. Bodenhausen, *J.Magn.Reson.Ser.A* **1993,** *102* 293-301.

[5.] E. Kupce, R. Freeman, *J.Magn.Reson.A* **1994,** *108* 268-273.

[6.] H. Geen, R. Freeman, *J.Magn.Reson.* **1991,** *93* 93-141.

[7.] A. J. Shaka, J. Keeler, R. Freeman, *J.Magn.Reson.* **1983,** *53* 313-340.

[8.] A. J. Shaka, P. B. Barker, R. Freeman, *J.Magn.Reson.* **1985,** *64* 547-552.

[9.] P. Andersson, J. Weigelt, G. Otting, *J.Biomol.NMR* **1998,** *12* 435-441.

[10.] M. Ottiger, F. Delaglio, A. Bax, *J.Magn.Reson.* **1998,** *131* 373-378.

[11.] L. Duma, S. Hediger, B. Brutscher, A. Böckmann, L. Emsley, *J.Am.Chem.Soc.* **2003,** *125* 11816-11817.

[12.] W. Bermel, I. Bertini, L. Duma, L. Emsley, I. C. Felli, R. Pierattelli, P. R. Vasos, *Angew.Chem.Int.Ed.* **2005,** *44* 3089-3092.

[13.] W. Bermel, I. C. Felli, R. Kümmerle, R. Pierattelli, *Concepts Magn.Reson.* **2008,** *32A* 183-200.

[14.] W. Bermel, I. Bertini, V. Csizmok, I. C. Felli, R. Pierattelli, P. Tompa, *J.Magn.Reson.* **2009,** *198* 275-281.

[15.] L. E. Kay, P. Keifer, T. Saarinen, *J.Am.Chem.Soc.* **1992,** *114* 10663-10665.

[16.] S. Grzesiek, A. Bax, *J.Am.Chem.Soc.* **1993,** *115* 12593-12594.

[17.] W. Bermel, I. Bertini, I. C. Felli, R. Kümmerle, R. Pierattelli, *J.Magn.Reson.* **2006,** *178* 56-64.

[18.] P. Schanda, B. Brutscher, *J.Am.Chem.Soc.* **2005,** *127* 8014-8015.

[19.] S. Mori, C. Abeygunawardana, M. O. Johnson, P. C. van Zijl, *J.Magn Reson.B* **1995,** *108* 94-98.

[20.] K. L. Lemay, J. J. Treadway, I. I. Angulo, T. L. Tellinghuisen, *J.Virol.* **2013,** *87* 1255-1260.

[21.] T. L. Tellinghuisen, K. L. Foss, J. J. Treadway, *PLoS.Pathog.* **2008,** *4* -e1000032.

5.1.2

Longitudinal relaxation properties of $^{1}H^{N}$ and $^{1}H^{\alpha}$ determined by ^{13}C direct-detected NMR experiments to study intrinsically disordered proteins (IDPs)

Tomáš Hošek[a], Sergio Gil-Gaballero[b], Roberta Pierattelli[a], Bernhard Brutscher[c], Isabella C. Felli[a]

[a] *CERM and Department of Chemistry Ugo Schiff, University of Florence, via Luigi Sacconi 6, Sesto Fiorentino, 50019, Italy*

[b] *Bruker BioSpin AG, Industriestrasse 26, Fällanden, 8117, Switzerland*

[c] *Institut de Biologie Structurale, Université Grenoble 1, CNRS, CEA, 71 avenue des Martyrs, 38044 Grenoble Cedex 9, France*

Submitted to *Journal of Magnetic Resonance*

Abstract

Intrinsically disordered proteins (IDPs) are functional proteins containing large fragments characterized by high local mobility. Bioinformatic studies have suggested that a significant fraction (more than 30%) of eukaryotic proteins has disordered regions of more than 50 amino acids in length. Hence, NMR methods for the characterization of local compactness and solvent accessibility in such highly disordered proteins are of high importance. Among the available approaches, the HET-SOFAST/BEST experiments (Schanda *et al.* 2006, Rennella *et al.* 2014) provide semi-quantitative information by monitoring longitudinal ^{1}H relaxation of amide protons under different initial conditions. However, when approaching physiological sample conditions, the potential of these amide ^{1}H detected experiments is reduced due to rapid amide proton solvent exchange. ^{13}C direct detection methods therefore provide a valuable alternative thanks to a higher chemical shift dispersion and their intrinsic insensitivity toward solvent exchange. Here we present two sets of ^{13}C detected experiments, which indirectly measure H^{N} and H^{α} inversion recovery profiles. The experiments consist of an initial spin inversion recovery block optimized for selective manipulation of different types of proton spins followed by a CON read-out scheme. The proposed experiments were tested on human α-synuclein and ubiquitin, two representative examples of unfolded and folded proteins.

Keywords

Longitudinal relaxation enhancement, ^{13}C direct detection, Intrinsically Disordered Proteins, IDPs

Introduction

Intrinsically disordered proteins (IDPs) are characterized by the lack of a unique stable three-dimensional structure under physiological conditions. [1–3] The high local mobility of these proteins provides advantages in terms of binding promiscuity, and allows them to interact with a large variety of binding partners. [4–8] This enables IDPs to play an important role in regulatory and signalling processes. Bioinformatic predictions show that a significant fraction of eukaryotic proteins is constituted by disordered regions of more than 50 amino acids in length. [9, 10] This has stimulated considerable interest in the development of novel NMR methods optimized for the study of IDPs or intrinsically disordered protein regions (IDPRs). [11–14]

The lack of a stable three-dimensional fold, the high flexibility, and the solvent exposed backbones typical of IDPs have a strong impact on the NMR observables. For highly flexible IDPs, the favorable transverse relaxation properties enable the design of complex, high dimensional NMR experiments with a large number of transfer steps and frequency editing periods. However, the poor chemical shift dispersion and the fast exchange rates of labile protons with the solvent pose challenges that need to be taken into account when designing NMR experiments. Several approaches have recently been proposed to overcome experimental limitations and to exploit at best the capabilities of NMR spectroscopy to characterize IDPs. For example, experiments of high dimensionality, exploiting non-uniform data sampling (NUS) strategies, longitudinal-relaxation enhancement (LRE), and either ^{1}H or ^{13}C direct detection, provide highly resolved NMR spectra. [14–31]These tools now enable the sequence specific assignment of large (up to about 400 residues) IDPs. [30, 32]

A main goal of the NMR investigation of an IDP is the characterization of the heterogeneous structural and dynamic properties throughout the polypeptide chain, often linked to its function. This is by no means a trivial task, due to the strong interplay between structural and dynamic contributions to the NMR observables, as illustrated be recent progress in the field. [13, 33–40] Therefore, new ways of accessing and describing the conformational heterogeneity of IDPs are required.

In this context, an interesting approach to access information about local compactness and solvent accessibility was proposed, based on the comparison of amide 1H longitudinal relaxation rates under different initial conditions. [41, 42] This HET-SOFAST/BEST approach was shown to be a robust and solid method to discriminate between folded and unfolded proteins, and to recognize protein regions of different local compactness and

solvent accessibility. However, when focusing on complex IDPs approaching physiological conditions, the low chemical shift dispersion in the amide ^{1}H dimension and solvent exchange-induced line-broadening of the peaks in 2D ^{1}H-^{15}N correlation spectra reduce the amount of atomic resolved information that can be obtained by this method.

Here we propose a complementary method based on ^{13}C detection to determine longitudinal relaxation properties of $^{1}H^{N}$ and $^{1}H^{\alpha}$ by measuring inversion recovery profiles under different initial conditions. Since proton spins are only used as a starting polarization source, the ^{13}C detected read-out schemes enable the detection of fast exchanging amide sites, and benefit from the good spectral resolution obtained for IDPs by exclusively heteronuclear NMR experiments. [43, 44] We show that these experiments provide information on the local structural and dynamic properties, in particular on local proton density, local correlation time, and solvent exchange rates. The experiments were tested on two well-studied proteins ubiquitin and α-synuclein as representative examples of a folded and of an unfolded protein.

Materials and methods

Samples

α-synuclein and ubiquitin were selected as representatives of an intrinsically disordered protein and of a structured protein, respectively.
Samples of 1.0 mM uniformly ^{13}C, ^{15}N labeled human α-synuclein in 20 mM phosphate buffer at pH 7.4 and 6.4 were prepared as previously described. [45] EDTA and NaCl were added to reach the final concentration of 0.5 and 200 mM, respectively, and 10% D_2O was added for the lock signal. A sample of 0.5 mM uniformly ^{13}C, ^{15}N labeled human ubiquitin prepared as previously described [46] in 20 mM HEPES buffer at pH 7 (10% D_2O was added for the lock signal).

NMR Experiments

NMR experiments were performed at 16.4 T on a Bruker Avance spectrometer operating at 700.06 MHz ^{1}H, 176.03 MHz ^{13}C and 70.9 MHz ^{15}N frequencies, equipped with a cryogenically cooled probehead optimized for ^{13}C direct detection.

The inversion recovery blocks, shown in Figure 1, were implemented before $H^{N\text{-flip}}CON$ and $H^{\alpha\text{-flip}}CON$ [47, 48] read-out schemes, respectively. The two reference pulse schemes are shown in the Supplementary Material (Figure S1 and S2). The common parameters

used for all experiments are described in detail in this paragraph, while those specific for each experiment are reported in the captions of the figures.

All pulses were applied with phase x, unless otherwise specified. In all the figures reported, band selective ^{13}C pulses are denoted by rounded shapes and rectangular symbols represent hard pulses. The wide and narrow symbols correspond to π and $\pi/2$ flip angle pulses. Pulse field gradients (G_Z line) are also indicated by rounded shapes. For ^{13}C band-selective $\pi/2$ and π pulses, Q5 (or time reversed Q5) and Q3 shapes [49] with durations of 300 and 220 μs, respectively, were used, except for the π pulses that should be band-selective on the C^{α} region (Q3, 860 μs) and for the adiabatic π pulse to invert C', C^{aro} and C^{ali} (smoothed Chirp 500 μs, 25% smoothing, 80 kHz sweep width, 11.3 kHz RF field strength). [50] The ^{1}H carrier was placed at 4.7 ppm, the ^{15}N carrier was placed at 116 ppm, or at 122.5 ppm to also include signals of proline residues, the ^{13}C band-selective pulses on C^{ali}, C^{α} and C' were given at the center of each region and the adiabatic pulse was adjusted to cover the entire ^{13}C region. Decoupling of ^{15}N was achieved with Garp-4 (1.0 kHz) sequences. [51] All experiments employ the IPAP approach to suppress the C^{α}-C' coupling in the direct acquisition dimension [43, 52, 53], where the in-phase (IP) and antiphase (AP) components were acquired and stored in an interleaved manner, doubling the number of FIDs recorded. The phase cycle, the method used for quadrature detection, the duration of the delays shown in the pulse sequences and the relative strengths of the gradients used (all gradients had sine shape and duration of 1 ms, unless otherwise specified) are reported case-by-case in the figure captions.

Carbon and nitrogen chemical shifts were referred indirectly to the ^{1}H standard using the conversion factor derived from the ratio of NMR frequencies. [54] Proton resonances were calibrated with respect to the signal of 2,2-dimethylsilapentane-5-sulfonic acid (DSS). The carbon signal of dioxane at 69.31 ppm was used as secondary reference for carbon chemical shifts.

Experimental data were acquired using Bruker TopSpin 1.3 software, processed with Bruker TopSpin 3.1 software and analysed with Bruker Dynamics Center.

Results and Discussion

^{13}C detected ^{1}H inversion recovery experiments

Two versions of the 2D $H^{flip}CON$ experiment, either starting with $^{1}H^{N}$ or with $^{1}H^{\alpha}$ magnetization, [47, 48] are reported in Figures S1 and S2. These pulse schemes provide the basic experiments that were modified to measure longitudinal relaxation of $^{1}H^{N}$ and $^{1}H^{\alpha}$, respec-

tively, by inserting an inversion-recovery (IR) building block before the start of the pulse sequence. Different IR blocks have been designed to invert all ^{1}H spins ("non-selective" experiment), or only a desired subset of ^{1}H spins (variants of "selective" experiments). Selective spin inversion can be achieved either through band selective pulses [55, 56], or by exploiting the coupling of ^{1}H spins to the directly attached heteronuclei (^{13}C and ^{15}N) during a spin evolution period of duration $1/J_{HX}$. [47, 48, 57] The second approach was chosen here as it can be easily implemented both for $^{1}H^{N}$ and $^{1}H^{\alpha}$. The various inversion sequences used in this study are shown in Figure 1, where the spin state of different pools of protons is indicated by colored arrows. Selection via the one-bond ^{1}H-^{15}N and ^{1}H-^{13}C scalar couplings allows to discriminate ^{15}N-bound spins, ^{13}C-bound spins, and all others (water ^{1}H, OH, ...). In addition, the selectivity of the ^{13}C 180° pulse determines whether only $^{1}H^{\alpha}$, $^{1}H^{\mathrm{aliphatic}}$, or all protons with the attached carbon ($^{1}H^{C}$) are inverted. By choosing different inversion blocks, IR profiles can be measured for different initial conditions in order to evaluate their effect on $^{1}H^{N}$ and $^{1}H^{\alpha}$ longitudinal relaxation.

^{1}H longitudinal relaxation mechanisms

The main spin interactions leading to ^{1}H longitudinal relaxation are dipole-dipole interactions between neighboring protons. [58, 59] The energy put into the system by radio-frequency pulses is transferred within the dipolar-coupled spin network in the protein (cross-relaxation or NOE effect) and is dissipated through the lattice (molecular motions). The efficiency of this relaxation mechanism depends on the effective local rotational tumbling correlation time, the local proton density, and on the state of the neighboring spins. Furthermore, the effective longitudinal relaxation of labile protons, which are localized in solvent-exposed regions is also significantly affected by exchange processes with water protons. Both of these two relaxation mechanisms couple different pools of protons, a situation that can be exploited for longitudinal relaxation enhancement (LRE) in proteins by selectively perturbing only one subset of spins, while leaving all others unperturbed. [14,15,60–65] Of course, other relaxation mechanisms also contribute to longitudinal ^{1}H relaxation (^{1}H chemical shift anisotropy and ^{1}H-^{13}C, ^{1}H-^{15}N dipole-dipole interactions). However, these additional interactions only contribute to a minor extent and, more importantly, do not couple different pools of proton spins, and thus will not introduce a differential effect depending on the initial states of different pools of proton spins. The main contributions to the effective longitudinal relaxation (dipolar cross-relaxation or hydrogen exchange) can be identified indirectly by measuring longitudinal ^{1}H relaxation under different initial conditions in which only selected subsets of protons are perturbed. [41,61]

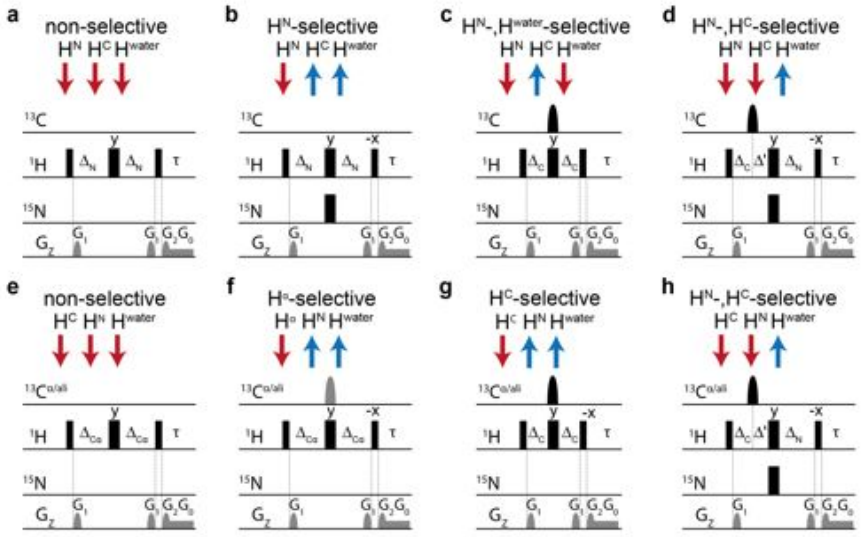


Figure 1. Inversion recovery blocks

Inversion recovery blocks implemented prior to the read-out pulse sequences. a) non-selective b) H^N-selective, c) H^N-H^{water}-selective and d) H^N-H^C-selective inversion recovery blocks implemented previously to the $H^{N\text{-flip}}$CON pulse sequence. e) non-selective f) H^α-selective, g) H^C-selective and h) H^N-H^C-selective inversion recovery blocks implemented prior to the $H^{\alpha\text{-flip}}$CON pulse sequence. The sets of ^{1}H spins inverted by each block are illustrated with red arrows. The following delays were used: $\Delta_N = 1/(2J_{NH})$, 5.2 ms; $\Delta_C = 1/(2J_{CH})$, 3.45 ms; $\Delta' = \Delta_N - \Delta_C$; $\Delta_{C\alpha} = 1/(2J_{C\alpha H\alpha})$, 3.75 ms. Adiabatic pulses on the carbon channel are indicated in black, highly selective pulses on $^{13}C^\alpha$ in grey. Gradient power relative to 53.5 G/cm was G_1:G_2:G_0 = 11:7:1%. A weak gradient (G_0) was applied during the inversion recovery delay to suppress radiation damping.

Two extreme cases can be imagined. Protons buried in the core of a structured protein are well protected from the solvent, and their longitudinal relaxation is determined by the dense ^{1}H-^{1}H dipolar coupling network (dipolar cross-relaxation or spin-diffusion mechanism). Since $^1H^N$ and $^1H^\alpha$ are very close in space it is likely that their relaxation properties will be very similar. In addition, longitudinal relaxation enhancement effects are expected to be similar for $^1H^N$ or $^1H^\alpha$ upon their selective perturbation with respect to all proton spins. [47,61] Of course, in this case it is expected that the state of the water polarization has no, or only a minor effect on the IR profiles. In the other extreme case of a proton in a largely disordered protein, both mechanisms will contribute and their relative importance will heavily depend on the temperature and pH of the sample. At

close to physiological conditions (high T, neutral pH), ^{1}H-^{1}H dipole-dipole interactions are expected to contribute only to a minor extent, while exchange effects are supposed to be very pronounced. This will translate into fast relaxation of $^1H^{\alpha}$ and moderate (or negligible) longitudinal relaxation enhancement effects, while exchangeable protons such as $^1H^N$ will show pronounced longitudinal relaxation enhancement effects. [14, 15]

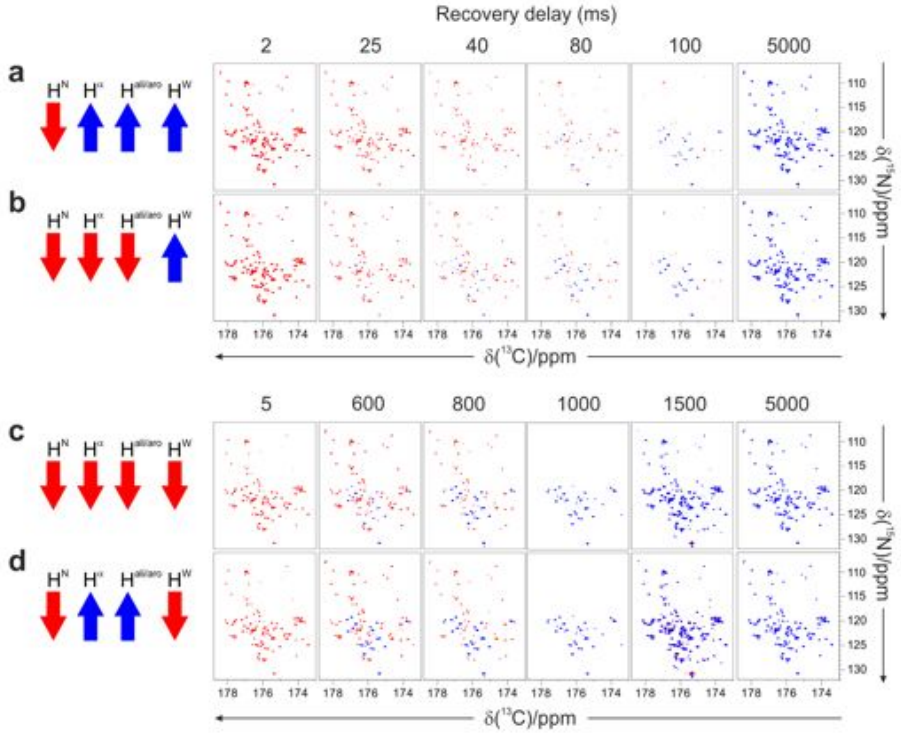


Figure 2. Series of IR-H$^{N\text{-}flip}$CON spectra of α-synuclein

An example of a series of the H$^{N\text{-}flip}$CON spectra measured with different inversion recovery blocks. The sets of ^{1}H spins inverted by the ^{1}H inversions blocks are illustrated with red arrows on the left: a) H^N-selective, b) H^N-H^C-selective, c) non-selective and d) H^N-H^{water}-selective. For the case a) and b) spectra with inversion recovery delays of 2 ms, 25 ms, 40 ms, 80 ms, 0.1 s and 5 s are shown, for the case c) and d) 5 ms, 0.6 s, 0.8 s, 1 s, 1.5 s and 5 s. Experiments were acquired on α-synuclein at pH 7.4 and 295.5 K (acquisition parameters are reported in Table S1 in Supplementary Material).

^{1}H^N and ^{1}H$^{\alpha}$ inversion recovery under different initial conditions

In order to access information on the relative importance of different mechanisms for ^{1}H longitudinal relaxation we have measured a series of IR-H$^{N\text{-}flip}$CON and IR-H$^{\alpha\text{-}flip}$CON

spectra using different inversion recovery schemes. Collecting data from both $^1H^N$ and $^1H^\alpha$ is expected to provide complementary information. The high chemical shift dispersion of 2D CON spectra [26] allowed us to determine IR profiles for the majority of $^1H^N$ and $^1H^\alpha$ protons in the proteins studied here.

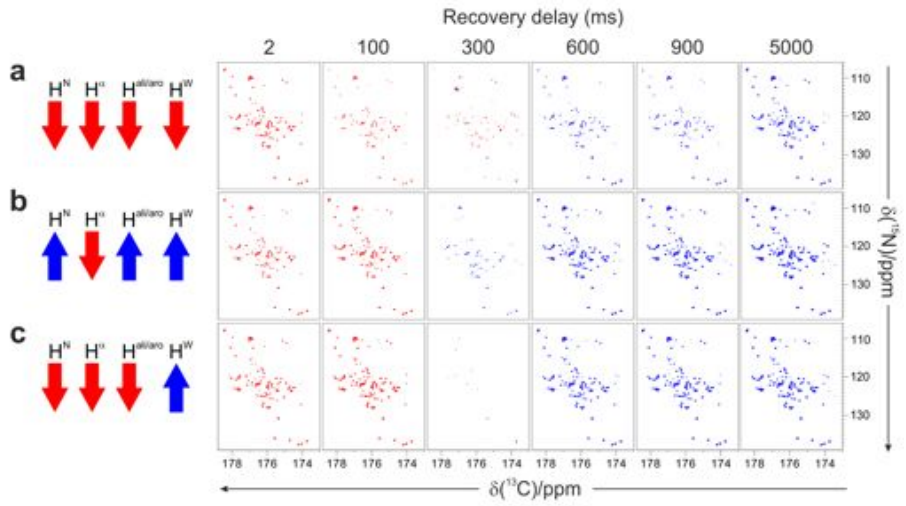


Figure 3. Series of IR-H$^{\alpha\text{-flip}}$CON spectra of α-synuclein
An example of a series of the H$^{\alpha\text{-flip}}$CON spectra measured with different inversion recovery blocks. The sets of ^1H spins inverted by the ^1H inversions blocks are illustrated with red arrows on the left: a) non-selective b) H^α-selective, c) H^N-H^C-selective. Spectra with inversion recovery delays of 2 ms, 0.1 s, 0.3 s, 0.6 s, 0.9 s and 5 s are shown. Experiments were acquired on α-synuclein at pH 7.4 and 295.5 K (acquisition parameters are reported in Table S1 in Supplementary Material).

Examples of IR-H$^{\text{N-flip}}$CON and IR-H$^{\alpha\text{-flip}}$CON spectra acquired on α-synuclein (295.5 K, pH 7.4) are shown in Figures 2 and 3, respectively. The zero-crossing of the IR curves ("T_{null}") can be used as an estimate of the effective ^1H T_1 (in case of mono-exponential spin recovery, $T_1 = 1.44 \cdot T_{null}$). In the case of $^1H^N$ longitudinal relaxation, a dramatic reduction of the T_{null} is observed, which changes from an average of about 1.2 s for non-selective inversion to about 50 ms for H^N-selective inversion. The data also show that the source of longitudinal relaxation enhancement comes from exchange processes with the bulk water rather than from dipole-dipole interactions, as the pronounced longitudinal relaxation enhancement effect observed is related to the state of the water polarization. For $^1H^\alpha$, T_{null} values of about 300 ms are measured, independent of the initial conditions that only marginally change the polarization recovery under the cho-

sen experimental conditions. As mentioned above, this is expected considering the high flexibility (cross-relaxation rates close to zero), the low proton density typical of highly disordered and flexible proteins, as well as the lack of direct exchange contributions.

Analogous IR profiles (Figure S3 and S4) acquired on the ubiquitin sample (295.5 K, pH 7) show a completely different behavior. For the majority of sites (residues), the $^1H^N$ and $^1H^\alpha$ IR profiles are quite similar, indicating the absence of significant exchange contributions and similar spin diffusion effects for the two neighboring protons. Only a few residues at the C-terminus, as well as in a highly flexible loop region connecting the first two β-strands show a different behavior, similar to the one observed for the majority of residues in α-synuclein.

Implications for the design of LRE experiments

The information obtained here about the longitudinal relaxation properties of H^N and H^α can be used to guide the design of new experiments and the choice of the optimal inter-scan delay. For example, if longitudinal relaxation enhancement effects are pronounced, it is advantageous to employ pulse sequences that actively exploit these effects to reduce the inter-scan delay and to increase the overall sensitivity. [14, 15, 31] On the other hand, if longitudinal relaxation enhancement effects are minor there is no need to implement any of the methods to achieve longitudinal relaxation enhancement. For experiments starting with amide ^1H polarization, this means that LRE will prove useful in almost any situation, independent of the sample conditions (pH and T) and whether the protein is structured or highly flexible. An exception may be the study of an IDP at high T and low pH, where exchange effects are negligible and cross-relaxation rates are close to zero. But this situation is of little practical relevance. In the case of H^α the situation is different: for well-structured proteins or protein regions LRE experiments will be of similar interest than for H^N-based experiments. The enhancement effect is of course only little dependent on the sample pH. For H^α sites in highly flexible protein regions, LRE provides an advantage in situations, e. g. low temperature, where the local rotational tumbling correlation times are slow enough to yield significant NOEs (and spin diffusion) with sequentially neighboring protons.

Information on conformational heterogeneity in proteins

The IR experiments presented here also provide information on the local rotational correlation time, proton density, and exchange processes. All these observables can be related

to local compactness, flexibility, and solvent accessibility along the polypeptide chain.

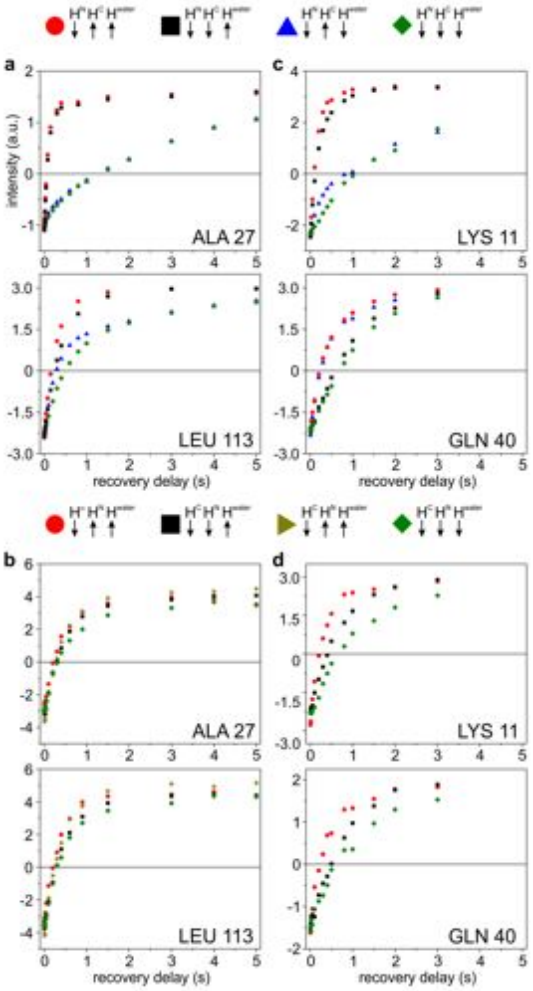


Figure 4. H^N and H^α inversion recovery profiles

Inversion recovery profiles of H^N (a, c) and H^α (b, d) acquired with the variant of the $H^{N\text{-flip}}CON$ and $H^{\alpha\text{-flip}}CON$, respectively, under different initial conditions as described in Figure 1. Inversion recovery profiles of alanine 27 and leucine 113 of α-synuclein (a, b), and lysine 11 and glutamine 40 of ubiquitin (c, d) are shown. Different initial conditions of H^N inversion recovery profiles (a, c) are indicated with the following symbols: (♦) non-selective, (▲) H^N-H^{water}-selective, (■) H^N-H^C-selective, and (●) H^N-selective. In the case of the H^α inversion recovery profiles (b, d) different initial conditions are indicated with the following symbols: (♦) non-selective, (►) H^C-selective, (■) H^N-H^C-selective, and (●) H^α-selective. The experiments were acquired at 295.5 K, the α-synuclein sample was at pH 7.5, the ubiquitin sample at pH 7.0.

Examples of inversion-recovery curves extracted from the 2D spectra of α-synuclein and ubiquitin are shown in Figure 4. These inversion-recovery curves illustrate distinct relaxation features observed for different protein regions. $^{1}H^{N}$ longitudinal relaxation of ALA 27, located in the amphiphilic N-terminal of α-synuclein, is strongly affected by the initial state of water spins, especially at conditions promoting rapid hydrogen exchange. LEU 113 that is part of the C-terminal region which has a high content of acidic residues, [66] displays a reduced dependence on the water ^{1}H spin state, even though the effect of hydrogen exchange is still significant. The differences in the $^{1}H^{\alpha}$ inversion recovery are minimal between the two sites. These observations agree with the intrinsically disordered nature of α-synuclein. They also point out that the charged C-terminal domain is electrostatically protected from the solvent, which decreases hydrogen exchange rates as previously described. [66–68]

In the case of human ubiquitin, the inversion recovery curves of one residue from a flexible loop region (LYS 11), and a second one from a well-structured part (GLN 40) are shown (Figure 4). The $^{1}H^{N}$ inversion recovery profile of LYS 11 is typical for solvent-exposed protein regions, where the main contribution to effective longitudinal relaxation comes from hydrogen exchange. Instead, the main relaxation mechanism of GLN 40 is ^{1}H-^{1}H spin-diffusion, since the initial spin state of the aliphatic and aromatic protons determines the recovery rate, whereas the effect of the water protons is minimal.

The longitudinal relaxation times obtained from the inversion recovery measurements under the various initial conditions tested in this study are plotted in Figures 5a and b for α-synuclein and in Figures 5c and d for ubiquitin as a function of the residue number. As discussed above, the observed dependence on the initial conditions reflects the different structural and dynamic properties of α-synuclein and ubiquitin. For ubiquitin most of the amino acid residues display similar effects. Only a highly flexible and solvent-exposed loop region, as well as the C-terminus can be easily distinguished from their pronounced relaxation enhancement effects observed for amide ^{1}H. In α-synuclein, the highly negatively charged segment at the C-terminus acts as a shielding factor for solvent exchange, resulting in smaller exchange rates in this segment. This observation could in principle be interpreted as a local structuring effect leading to a more compact conformation in this part of the intrinsically disordered protein. However, the small variations in the observed $^{1}H^{\alpha}$ relaxation data do provide valuable complementary information, clearly pointing towards the conclusion that significant structuring in this region does not occur.

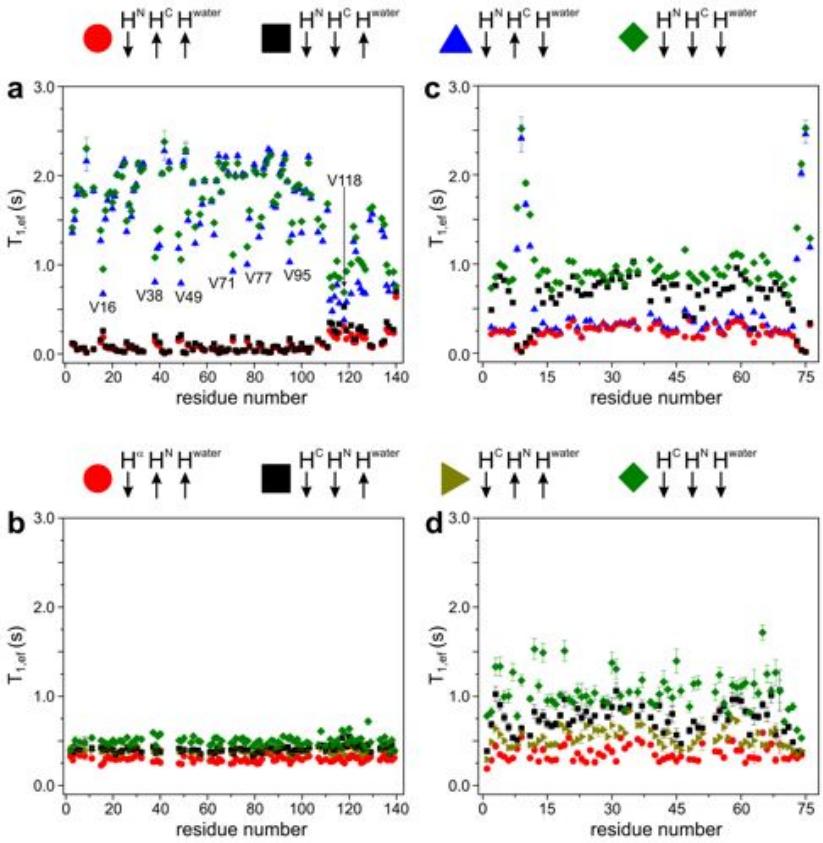


Figure 5. Estimates of effective longitudinal relaxation constants
Estimates of $^1H^N$ and $^1H^\alpha$ effective longitudinal relaxation time constants (effective T_1) for each residue of α-synuclein (a, b) and ubiquitin (c, d) was obtained by fitting inversion recovery data to a mono-exponential function. For H^N T_1 values (a,c), different initial conditions are indicated by the following symbols: (♦) non-selective, (▲) H^N-H^{water}-selective, (■) H^N-H^C-selective, and (●) H^N-selective. For H T1 values values (b,d), different initial conditions are indicated by: (♦) non-selective, (►) H^C-selective, (■) H^N-H^C-selective, and (●) H^α-selective. The experiments were acquired at 295.5 K, the α-synuclein sample was at pH 7.5, the ubiquitin sample at pH 7.0.

Similarly, care should be taken when attempting to translate the measured hydrogen-exchange induced relaxation enhancement into local protection factors indicating the degree of local compactness, as the chemical nature of the local amino acid composition has a strong impact on the observed effects. For example, all valine residues in α-synuclein experience more moderate enhancement effects, due to their intrinsically smaller exchange rates related to their specific chemical topology.

Conclusions

The high local mobility and fast solvent-exchange, inherent properties of IDPs, complicate their residue-specific structural and dynamic characterization. Here we have described a set of proton-selective inversion recovery blocks, that combined with ^{13}C-detected NMR experiments allowed us to measure effective proton longitudinal relaxation rates under different initial conditions. The proposed ^{13}C detected experiments were successfully tested on two model proteins, human α-synuclein and ubiquitin.

Our results, first of all, provide information about the overall longitudinal ^{1}H relaxation properties of the two proteins, and the main longitudinal relaxation mechanisms that can be exploited for longitudinal relaxation enhancement in multidimensional NMR experiments. They can also guide the optimal choice of the most appropriate pulse scheme and the setting of the inter-scan delay to achieve maximal sensitivity.

In addition, some useful information on the local structural and dynamic properties of the protein can be obtained under physiological sample conditions, e.g. at neutral pH and high temperature or under in-cell conditions, where IDPs are difficult to study by conventional NMR methods, such as ^{15}N relaxation measurements.

Acknowledgments

This work has been supported in part by the European Commission Projects IDPbyNMR (contract no. 264257), BioNMR (contract no. 261863) and INSTRUCT (contract no. 211252). Leonardo Gonnelli is acknowleged for providing the α-synuclein samples. Alessandro Piai and Zsófia Sólyom are gratefully acknowledged for stimulating discussions on the work and on the manuscript.

References

[1] Uversky VN, Gillespie JR, Fink AL. Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins*. 2000;41(3):415–427.

[2] Dunker AK, Lawson JD, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CR, Hipps KW, Ausio J, Nissen MS, Reeves R, Kang CH, Kissinger CR, Bailey RW, Griswold MD, Chiu M, Garner EC, Obradovic Z. Intrinsically disordered protein. In: *J Mol Graph Model*. Washington State Univ, Sch Mol Biosci, Pullman, WA 99164 USA. 2001; pp. 26–59.

[3] Tompa P, Fersht A. *Structure and Function of Intrinsically Disordered Proteins*. CRC Press. 2009.

[4] Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN. Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS Journal*. 2005; 272(20):5129–5148.

[5] Tompa P, Fuxreiter M. Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends in Biochemical Sciences*. 2008;33(1):2–8.

[6] Wright PE, Dyson HJ. Linking folding and binding. *Current Opinion in Structural Biology*. 2009;19(1):31–38.

[7] Mittag T, Marsh J, Grishaev A, Orlicky S, Lin H, Sicheri F, Tyers M, Forman-Kay JD. Structure/Function Implications in a Dynamic Complex of the Intrinsically Disordered Sic1 with the Cdc4 Subunit of an SCF Ubiquitin Ligase. *Structure*. 2010; 18(4):494–506.

[8] Habchi J, Tompa P, Longhi S, Uversky VN. Introducing Protein Intrinsic Disorder. *Chem Rev*. 2014;114(13):6561–6588.

[9] Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradovic Z. Intrinsic disorder and protein function. *Biochemistry*. 2002;41(21):6573–6582.

[10] Tompa P. Unstructural biology coming of age. *Current Opinion in Structural Biology*. 2011;21(3):419–425.

[11] Felli IC, Pierattelli R. Novel methods based on 13C detection to study intrinsically disordered proteins. *Journal of Magnetic Resonance*. 2014;241(C):115–125.

[12] Nováček J, Žídek L, Sklenář V. Toward optimal-resolution NMR of intrinsically disordered proteins. *Journal of Magnetic Resonance*. 2014;241:41–52.

[13] Konrat R. NMR contributions to structural dynamics studies of intrinsically disordered proteins . *Journal of Magnetic Resonance*. 2014;241:74–85.

[14] Solyom Z, Schwarten M, Geist L, Konrat R, Willbold D, Brutscher B. BEST-TROSY experiments for time-efficient sequential resonance assignment of large disordered proteins. *J Biomol NMR*. 2013;55(4):311–321.

[15] Gil S, Hošek T, Solyom Z, Kümmerle R, Brutscher B, Pierattelli R, Felli IC. NMR Spectroscopic Studies of Intrinsically Disordered Proteins at Near-Physiological Conditions. *Angew Chem Int Ed.* 2013;52(45):11808–11812.

[16] Bermel W, Bertini I, Felli IC, Gonnelli L, Koźmiński W, Piai A, Pierattelli R, Stanek J. Speeding up sequence specific assignment of IDPs. *J Biomol NMR.* 2012;53(4):293–301.

[17] Bermel W, Felli IC, Gonnelli L, Koźmiński W, Piai A, Pierattelli R, Zawadzka-Kazimierczuk A. High-dimensionality C-13 direct-detected NMR experiments for the automatic assignment of intrinsically disordered proteins. *J Biomol NMR.* 2013; 57(4):353–361.

[18] Narayanan RL, Dürr UHN, Bibow S, Biernat J, Mandelkow E, Zweckstetter M. Automatic Assignment of the Intrinsically Disordered Protein Tau with 441-Residues. *J Am Chem Soc.* 2010;132(34):11906–11907.

[19] Fiorito F, Hiller S, Wider G, Wüthrich K. Automated Resonance Assignment of Proteins: 6 DAPSY-NMR. *J Biomol NMR.* 2006;35(1):27–37.

[20] Chakraborty S, Susmitha AL, Hosur RV. Selective lighting up of segments around Gly, Ala and Ser/Thr in proteins. *Magn Reson Chem.* 2012;50(9):587–591.

[21] Kumar D, Hosur RV. hNCOcanH pulse sequence and a robust protocol for rapid and unambiguous assignment of backbone (H-1(N), N-15 and C-13 ') resonances in N-15/C-13-labeled proteins. *Magn Reson Chem.* 2011;49(9):575–583.

[22] Pantoja-Uceda D, Santoro J. New 13C-detected experiments for the assignment of intrinsically disordered proteins. *J Biomol NMR.* 2014;59(1):43–50.

[23] Pantoja-Uceda D, Santoro J. Direct correlation of consecutive C'-N groups in proteins: a method for the assignment of intrinsically disordered proteins. *J Biomol NMR.* 2013; 57(1):57–63.

[24] Mäntylahti S, Aitio O, Hellman M, Permi P. HA-detected experiments for the backbone assignment of intrinsically disordered proteins. *J Biomol NMR.* 2010;47(3):171–181.

[25] Bermel W, Bertini I, Csizmok V, Felli IC, Pierattelli R, Tompa P. H-start for exclusively heteronuclear NMR spectroscopy: The case of intrinsically disordered proteins. *Journal of Magnetic Resonance.* 2009;198(2):275–281.

[26] Bermel W, Bertini I, Felli IC, Lee YM, Luchinat C, Pierattelli R. Protonless NMR Experiments for Sequence-Specific Assignment of Backbone Nuclei in Unfolded Proteins. *J Am Chem Soc.* 2006;128(12):3918–3919.

[27] Kazimierczuk K, Stanek J, Zawadzka-Kazimierczuk A, Koźmiński W. High-dimensional NMR spectra for structural studies of biomolecules. *ChemPhysChem.* 2013;14(13):3015–3025.

[28] Nováček J, Zawadzka-Kazimierczuk A, Papoušková V, Žídek L, Šanderová H, Krásný L, Koźmiński W, Sklenář V. 5D 13C-detected experiments for backbone assignment of unstructured proteins with a very low signal dispersion. *J Biomol NMR.* 2011; 50(1):1–11.

[29] Nováček J, Haba NY, Chill JH, Žídek L, Sklenář V. 4D non-uniformly sampled HCBCACON and J(NCα)-selective HCBCANCO experiments for the sequential assignment and chemical shift analysis of intrinsically disordered proteins. *J Biomol NMR.* 2012;53(2):139–148.

[30] Nováček J, Janda L, Dopitová R, Žídek L, Sklenář V. Efficient protocol for backbone and side-chain assignments of large, intrinsically disordered proteins: transient secondary structure analysis of 49.2 kDa microtubule associated protein 2c. *J Biomol NMR.* 2013;56(4):291–301.

[31] Piai A, Hošek T, Gonnelli L, Zawadzka-Kazimierczuk A, Koźmiński W, Brutscher B, Bermel W, Pierattelli R, Felli IC. "CON-CON" assignment strategy for highly flexible intrinsically disordered proteins. *J Biomol NMR.* 2014;60(4):209–218.

[32] Mukrasch MD, Bibow S, Korukottu J, Jeganathan S, Biernat J, Griesinger C, Mandelkow E, Zweckstetter M. Structural polymorphism of 441-residue tau at single residue resolution. *PLoS Biol.* 2009;7(2):e34.

[33] Parigi G, Rezaei-Ghaleh N, Giachetti A, Becker S, Fernandez C, Blackledge M, Griesinger C, Zweckstetter M, Luchinat C. Long-Range Correlated Dynamics in Intrinsically Disordered Proteins. *J Am Chem Soc.* 2014;136:16201–16209.

[34] Kragelj J, Ozenne V, Blackledge M, Jensen MR. Conformational Propensities of Intrinsically Disordered Proteins from NMR Chemical Shifts. *ChemPhysChem.* 2013; 14(13):3034–3045.

[35] Camilloni C, De Simone A, Vranken WF, Vendruscolo M. Determination of Secondary Structure Populations in Disordered States of Proteins Using Nuclear Magnetic Resonance Chemical Shifts. *Biochemistry*. 2012;51(11):2224–2231.

[36] Jensen MR, Markwick PRL, Meier S, Griesinger C, Zweckstetter M, Grzesiek S, Bernadó P, Blackledge M. Quantitative Determination of the Conformational Properties of Partially Folded and Intrinsically Disordered Proteins Using NMR Dipolar Couplings. *Structure/Folding and Design*. 2009;17(9):1169–1185.

[37] Stanek J, Saxena S, Geist L, Konrat R, Koźmiński W. Probing local backbone geometries in intrinsically disordered proteins by cross-correlated NMR relaxation. *Angew Chem Int Ed Engl*. 2013;52(17):4604–4606.

[38] Camilloni C, Vendruscolo M. Statistical mechanics of the denatured state of a protein using replica-averaged metadynamics. *J Am Chem Soc*. 2014;136(25):8982–8991.

[39] Forman-Kay JD, Mittag T. From Sequence and Forces to Structure, Function, and Evolution of Intrinsically Disordered Proteins. *Structure/Folding and Design*. 2013; 21(9):1492–1499.

[40] Jensen MR, Zweckstetter M, Huang JR, Blackledge M. Exploring free-energy landscapes of intrinsically disordered proteins at atomic resolution using NMR spectroscopy. *Chem Rev*. 2014;114(13):6632–6660.

[41] Schanda P, Forge V, Brutscher B. HET-SOFAST NMR for fast detection of structural compactness and heterogeneity along polypeptide chains. *Magn Reson Chem*. 2006; 44(S1):S177–S184.

[42] Rennella E, Solyom Z, Brutscher B. Measuring hydrogen exchange in proteins by selective water saturation in (1)H- (15)N SOFAST/BEST-type experiments: advantages and limitations. *J Biomol NMR*. 2014;60(2-3):99–107.

[43] Bermel W, Bertini I, Felli IC, Piccioli M, Pierattelli R. 13C-detected protonless NMR spectroscopy of proteins in solution. *Progress in Nuclear Magnetic Resonance Spectroscopy*. 2006;48(1):25–45.

[44] Felli IC, Piai A, Pierattelli R. Recent Advances in Solution NMR Studies: 13C Direct Detection for Biomolecular NMR Applications. *Ann Rep NMR Spectroscop*. 2013;(6):359–418.

[45] Huang CJ, Ren GP, Zhou H, Wang CC. A new method for purification of recombinant human alpha-synuclein in Escherichia coli. *Protein Expression and Purification*. 2005; 42(1):173–177.

[46] Wand AJ, Urbauer JL, McEvoy RP, Bieber RJ. Internal Dynamics of Human Ubiquitin Revealed by 13C-Relaxation Studies of Randomly Fractionally Labeled Protein . *Biochemistry*. 1996;35(19):6116–6125.

[47] Bermel W, Bertini I, Felli IC, Pierattelli R. Speeding Up 13C Direct Detection Biomolecular NMR Spectroscopy. *J Am Chem Soc*. 2009;131(42):15339–15345.

[48] Bertini I, Felli IC, Gonnelli L, Kumar V M V, Pierattelli R. High-Resolution Characterization of Intrinsic Disorder in Proteins: Expanding the Suite of 13C-Detected NMR Spectroscopy Experiments to Determine Key Observables. *ChemBioChem*. 2011;12(15):2347–2352.

[49] Emsley L, Bodenhausen G. Optimization of Shaped Selective Pulses for Nmr Using a Quaternion Description of Their Overall Propagators. *J Magn Reson*. 1992;97(1):135–148.

[50] Bohlen JM, Bodenhausen G. Experimental Aspects of Chirp Nmr-Spectroscopy. *J Magn Reson A*. 1993;102(3):293–301.

[51] Shaka AJ, Barker PB, Freeman R. Computer-optimized decoupling scheme for wideband applications and low-level operation. *J Magn Reson (1969-1992)*. 1985; 64(3):547–552.

[52] Bermel W, Bertini I, Felli IC, Kümmerle R, Pierattelli R. Novel 13C direct detection experiments, including extension to the third dimension, to perform the complete assignment of proteins. *Journal of Magnetic Resonance*. 2006;178(1):56–64.

[53] Bermel W, Felli IC, Kümmerle R, Pierattelli R. 13C Direct-detection biomolecular NMR. *Concepts Magn Reson*. 2008;32A(3):183–200.

[54] Markley JL, Bax A, Arata Y, Hilbers CW, Kaptein R, Sykes BD, Wright PE, Wüthrich K. Recommendations for the presentation of NMR structures of proteins and nucleic acids–IUPAC-IUBMB-IUPAB Inter-Union Task Group on the standardization of data bases of protein and nucleic acid structures determined by NMR spectroscopy. *Eur J Biochem*. 1998;256(1):1–15.

[55] Schanda P, Brutscher B. Very Fast Two-Dimensional NMR Spectroscopy for Real-Time Investigation of Dynamic Events in Proteins on the Time Scale of Seconds. *J Am Chem Soc.* 2005;127(22):8014–8015.

[56] Lescop E, Kern T, Brutscher B. Guidelines for the use of band-selective radiofrequency pulses in hetero-nuclear NMR: example of longitudinal-relaxation-enhanced BEST-type 1H-15N correlation experiments. *J Magn Reson.* 2010;203(1):190–198.

[57] Diercks T, Daniels M, Kaptein R. Extended flip-back schemes for sensitivity enhancement in multidimensional HSQC-type out-and-back experiments. *J Biomol NMR.* 2005;33(4):243–259.

[58] Levitt MH. *Spin Dynamics.* Basics of Nuclear Magnetic Resonance. Wiley. 2008.

[59] Cavanagh J, Fairbrother WJ, Palmer AG III, Skelton NJ, Rance M. *Protein NMR Spectroscopy, Second Edition: Principles and Practice.* San Diego: Academic Press, 2nd ed. 2006.

[60] Schanda P, Kupče E, Brutscher B. SOFAST-HMQC Experiments for Recording Two-dimensional Deteronuclear Correlation Spectra of Proteins within a Few Seconds. *J Biomol NMR.* 2005;33(4):199–211.

[61] Schanda P. Fast-pulsing longitudinal relaxation optimized techniques: Enriching the toolbox of fast biomolecular NMR spectroscopy. *Progress in Nuclear Magnetic Resonance Spectroscopy.* 2009;55(3):238–265.

[62] Felli IC, Brutscher B. Recent Advances in Solution NMR: Fast Methods and Heteronuclear Direct Detection. *ChemPhysChem.* 2009;10(9-10):1356–1368.

[63] Pervushin K, Vogeli B, Eletsky A. Longitudinal H-1 relaxation optimization in TROSY NMR spectroscopy. *J Am Chem Soc.* 2002;124(43):12898–12902.

[64] Banci L, Luchinat C. Selective versus non-selective T-1 experiments to determine metal-nucleus distances in paramagnetic metalloproteins. *Inorganica Chimica Acta.* 1998;276(1-2):373–379.

[65] Deschamps M, Campbell ID. Cooling overall spin temperature: protein NMR experiments optimized for longitudinal relaxation effects. *J Magn Reson.* 2006;178(2):206–211.

[66] Croke RL, Sallum CO, Watson E, Watt ED, Alexandrescu AT. Hydrogen exchange of monomeric α-synuclein shows unfolded structure persists at physiological temperature and is independent of molecular crowding in Escherichia coli. *Protein Science*. 2008; 17(8):1434–1445.

[67] Kim PS, Baldwin RL. Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. *Annu Rev Biochem*. 1982;51(1):459–489.

[68] Matthew JB, Weber PC, Salemme FR, Richards FM. Electrostatic orientation during electron transfer between flavodoxin and cytochrome c. *Nature*. 1983;301(5896):169–171.

Longitudinal relaxation properties of $^{1}H^{N}$ and $^{1}H^{\alpha}$ determined by ^{13}C direct-detected NMR experiments to study intrinsically disordered proteins (IDPs)

Tomáš Hošek[a], Sergio Gil-Gaballero[b], Roberta Pierattelli[a], Bernhard Brutscher[c], Isabella C. Felli[a]

[a] *CERM and Department of Chemistry Ugo Schiff, University of Florence, via Luigi Sacconi 6, Sesto Fiorentino, 50019, Italy*

[b] *Bruker BioSpin AG, Industriestrasse 26, Fällanden, 8117, Switzerland*

[c] *Institut de Biologie Structurale, Université Grenoble 1, CNRS, CEA, 71 avenue des Martyrs, 38044 Grenoble Cedex 9, France*

SUPPLEMENTRY MATERIAL

Table S1: The NMR acquisition parameters used for inversion recovery (IR) $H^{N\text{-flip}}CON$ and $H^{\alpha\text{-flip}}CON$ experiments

Experiments	**Dimension of acquired data**		**Spectral width (ppm)**		n[a]	d[b]
	t_1	t_2	F_1	F_2		
IR-$H^{N\text{-flip}}$CON-IPAP[c]						
α-synuclein (pH 7.4, 295.5 K)[d]	256*(^{15}N)	1024(^{13}C)	35	30	2	5
α-synuclein (pH 6.4, 295.5 K)[d]	256*(^{15}N)	1024(^{13}C)	35	30	2	5
α-synuclein (pH 6.4, 285.7 K)[d]	256*(^{15}N)	1024(^{13}C)	35	30	2	5
ubiquitin (pH 7.0, 295.5 K)[e]	180*(^{15}N)	1024(^{13}C)	30	30	2	5
ubiquitin (pH 7.0, 285.7 K)[e]	180*(^{15}N)	1024(^{13}C)	30	30	2	5
IR-$H^{\alpha\text{-flip}}$CON-IPAP[c]						
α-synuclein (pH 7.4, 295.5 K)[f]	400*(^{15}N)	1024(^{13}C)	35	30	2	5
α-synuclein (pH 6.4, 295.5 K)[f]	256*(^{15}N)	1024(^{13}C)	35	30	2	5
ubiquitin (pH 7.0, 295.5 K)[e]	180*(^{15}N)	1024(^{13}C)	30	30	8	5
ubiquitin (pH 7.0, 290.6 K)[e]	180*(^{15}N)	1024(^{13}C)	30	30	8	5

[a] Number of acquired scans.

[b] Relaxation delay in seconds.

[c] For experiments acquired in the IPAP mode, the dimension in which the two experiments are stored is indicated with an asterisk

[d] Inversion recovery delays were 5, 15, 25, 50 ms, 0.1, 0.2, 0.4, 0.8, 1, 1.5, 2, 3, 5 s for non-selective and H^N-H^{water}-selective case, and 5, 8, 10, 15, 25, 40, 80 ms, 0.15, 0.3, 0.4, 0.8, 2, 5 s for H^N-selective and H^N-H^C-selective case.

[e] Inversion recovery delays were 5, 10, 25, 50 ms, 0.1, 0.2, 0.3, 0.4, 0.5, 0.8, 1, 1.5, 2, 3 s in all cases.

[f] Inversion recovery delays were 2, 5, 10, 15, 25, 50 ms, 0.1, 0.2, 0.3, 0.4, 0.6, 0.9, 1.5, 3, 4, 5 s in all cases.

Figure S1. The $H^{N\text{-flip}}CON$ pulse sequence. The delays are: $\Delta_1 = 1/(2J_{NH})$, 5.4 ms; $\Delta_2 = 1/(2J_{C'N})$, 33.2 ms; $\Delta_3 = 1/(2J_{C'C\alpha})$, 9 ms; $\Delta = 20$ ms to increase the resolution in the ^{15}N dimension. The phase cycle is ϕ_1 = x,-x; ϕ_2 = -x,x; ϕ_3 = y,y,-y,-y; ϕ_{rec} = x,-x,x,-x. Quadrature detection was obtained by incrementing phase ϕ_3 (t_1) in States-TPPI manner. Gradient power relative to 53.5 G/cm was G_1:G_2:G_3:G_4 = 5:17:33:50%.

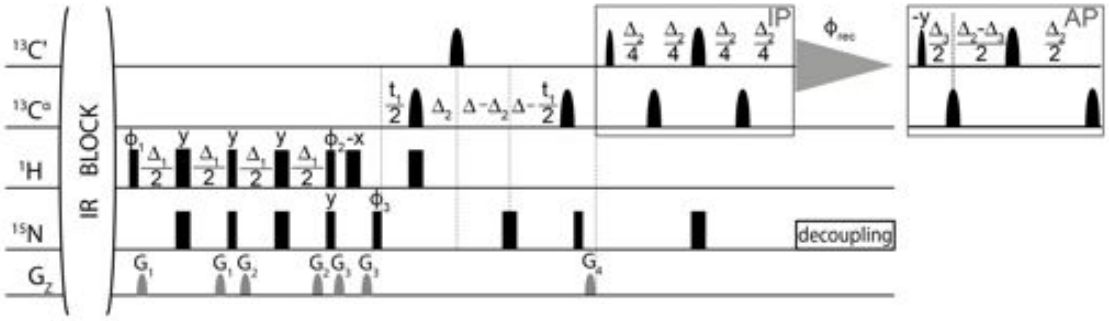


Figure S2. The $H^{\alpha\text{-flip}}CON$ pulse sequence. The delays are: $\Delta_1 = 1/(3J_{HC\alpha})$, 2.7 ms; $\Delta_1' = 1/(6J_{HC\alpha})$, 1.1 ms; $\Delta_2 = 1/(2J_{C'C\alpha})$, 9 ms; $\Delta_3 = 1/(J_{C\alpha C\beta})$, 24 ms; $\Delta_4 = 1/(2J_{C'N})$, 33.2 ms, $\Delta = t_1(0) + p8$, where p8 is the duration of the adiabatic pulse on ^{13}C in the middle of the ^{15}N evolution period. The phase cycle is ϕ_1 = x,x,-x,-x; ϕ_2 = -x,x; ϕ_{rec} = x,-x,-x,x. Quadrature detection was obtained by incrementing phase ϕ_1 (t_1) in States-TPPI manner. Gradient power relative to 53.5 G/cm was G_1:G_2:G_3:G_4:G_5 = 5:17:33:50:-20%. Highly selective pulses on C^α are indicated in grey. Adiabatic pulse on the carbons is hatched. The pulse indicated with asterisk is applied in the centre of $C^{aliphatic}$ region.

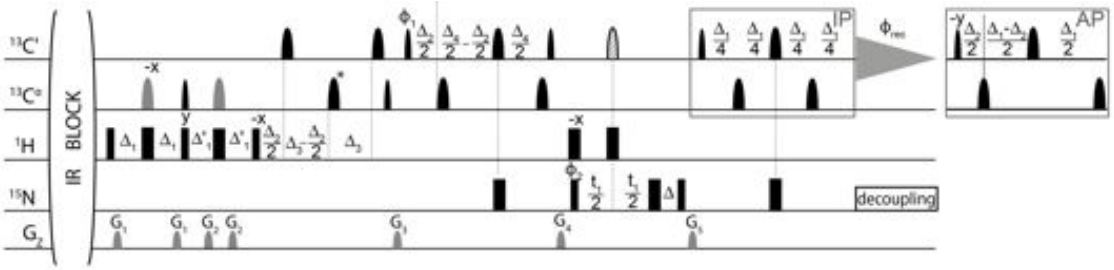


Figure S3. An example of a series of the $H^{N\text{-flip}}CON$ spectra with implemented inversion recovery blocks. The sets of ^{1}H spins inverted by the blocks are illustrated with red arrows on the left: a) H^{N}-selective, b) H^{N}-H^{C}-selective, c) non-selective and d) H^{N}-H^{water}-selective. The spectra with inversion recovery delays of 5, 50 ms, 0.2, 0.4, 0.8 and 3 s are shown. Experiments were acquired on human ubiquitin sample at pH 7.0 and 295.5 K (acquisition parameters are reported in Table S1)

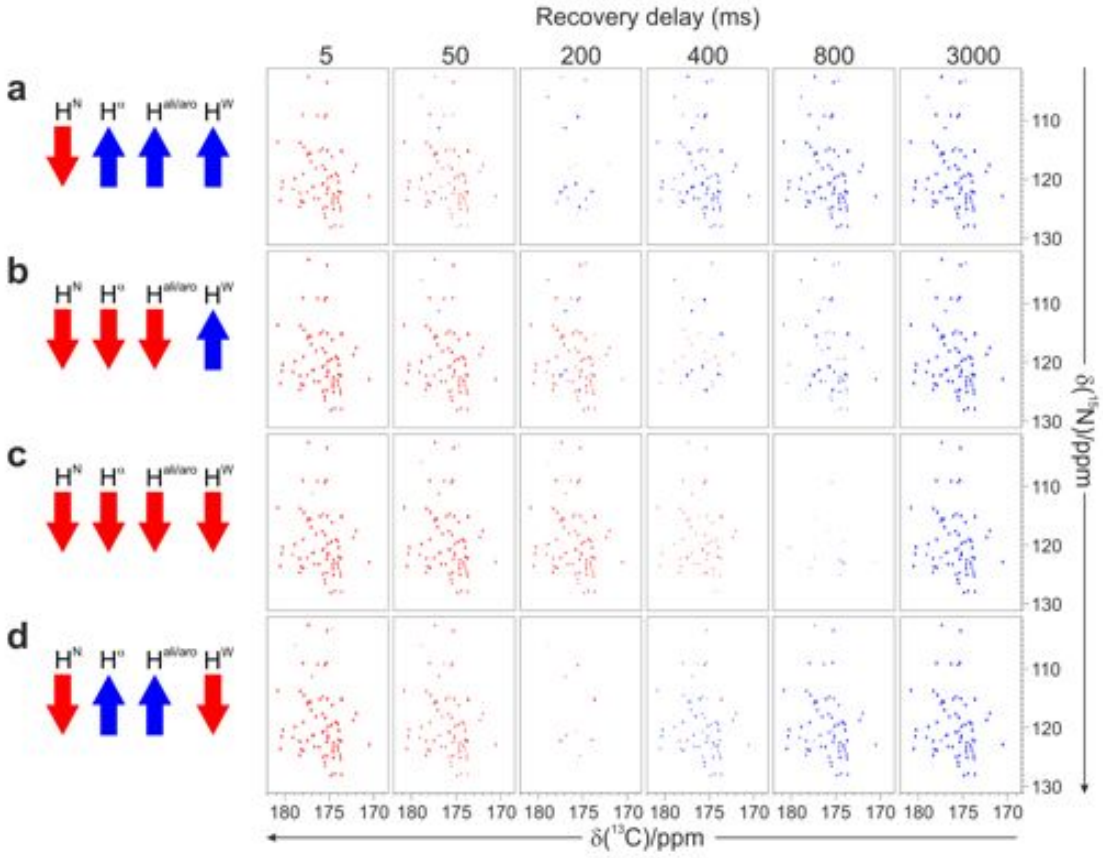


Figure S4. An example of a series of the $H^{\alpha\text{-flip}}$CON spectra with implemented inversion recovery blocks. The sets of ^{1}H spins inverted by the blocks are illustrated with red arrows on the left: a) non-selective, b) H^{α}-selective and c) H^{N}-H^{C}-selective. Spectra with inversion recovery delays 5, 50 ms, 0.2, 0.4, 0.8 and 3 s are shown. Experiments were acquired on human ubiquitin sample at pH 7.0 and 295.5 K (acquisition parameters are reported in Table S1)

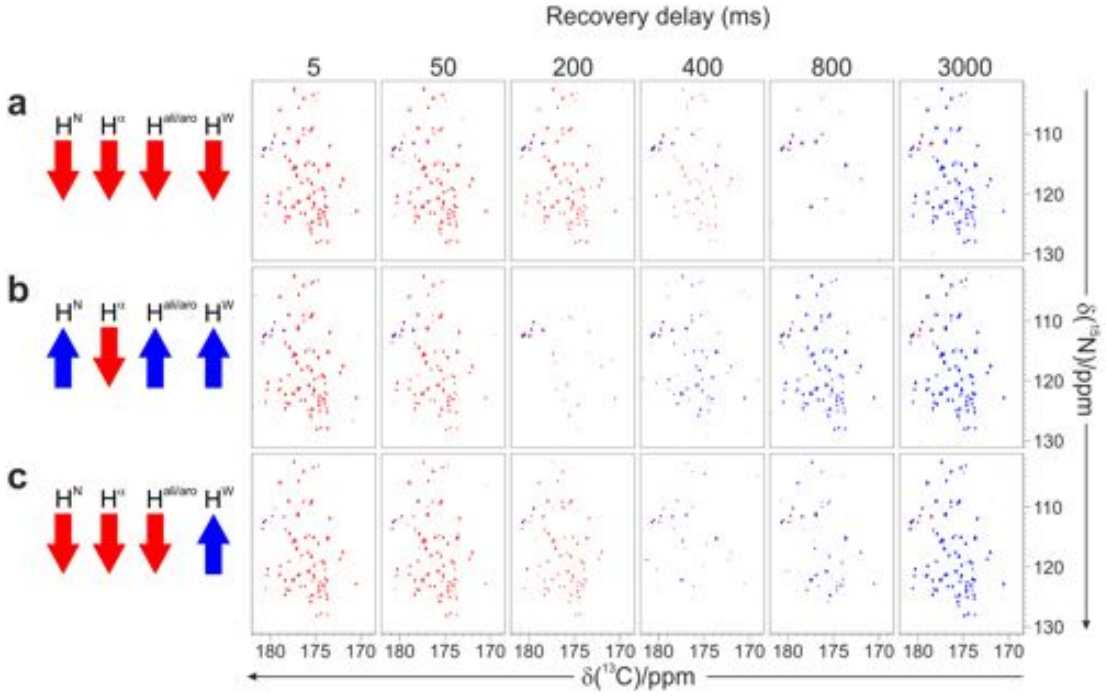


Figure S5. Inversion recovery profiles of H^N and H^α acquired with the variant of the $H^{N\text{-flip}}$CON and $H^{\alpha\text{-flip}}$CON, respectively, under different initial conditions as described in Figure 1. Inversion recovery profiles of selected amino acids of α-synuclein are shown: a,c) Alanine 27 and b,d) Leucine 113. Different initial conditions of H^N inversion recovery profiles (a,b) are indicated with the following symbols: (♦) non-selective, (▲) H^N-H^{water}-selective, (■) H^N-H^C-selective, and (●) H^N-selective. In the case of the H^α inversion recovery profiles (c,d) different initial conditions are indicated with the following symbols: (♦) non-selective, (►) H^C-selective, (■) H^N-H^C-selective, and (●) H^α-selective. The pH of the sample, and temperature at which experiments were conducted are specified in each graph.

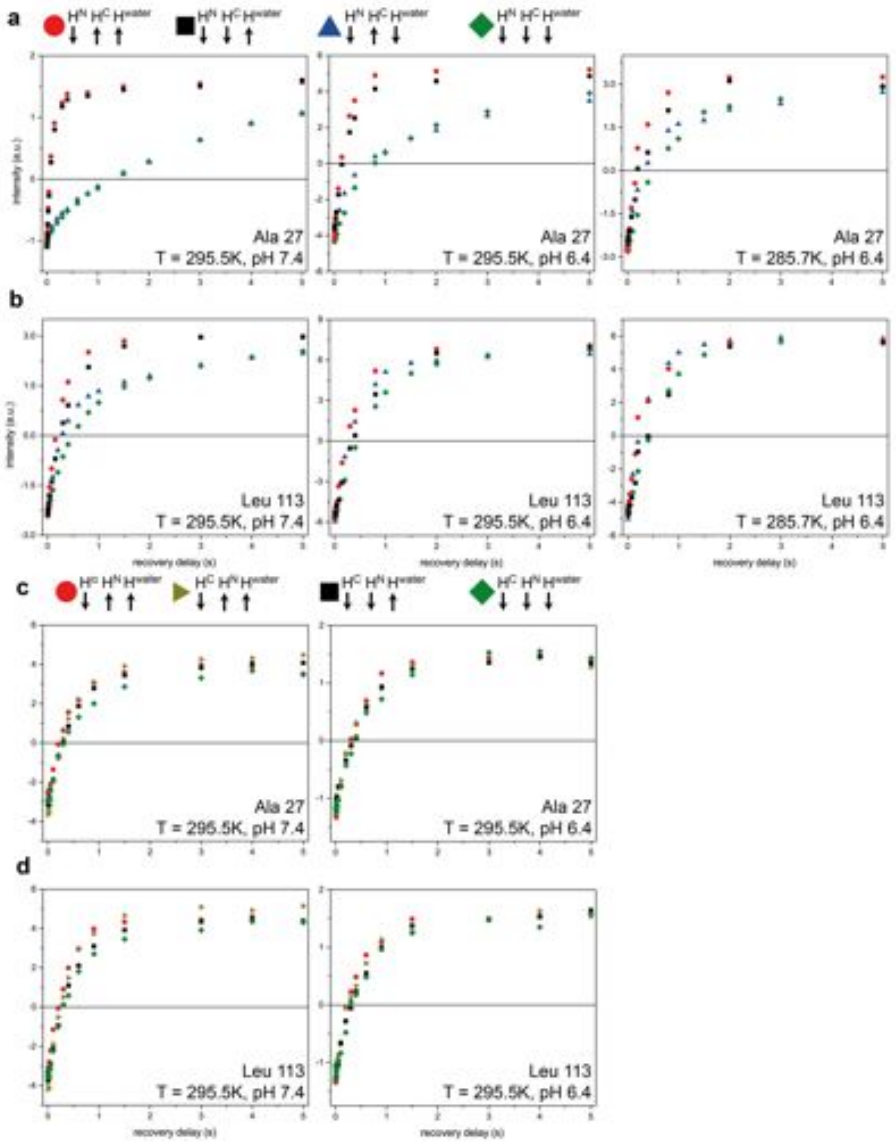


Figure S6. Inversion recovery profiles of H^N and H^α acquired with the variant of the $H^{N\text{-flip}}CON$ and $H^{\alpha\text{-flip}}CON$, respectively, under different initial conditions as described in Figure 1. Inversion recovery profiles of selected amino acids of ubiquitin are shown: a,c) Lysine 11 and b,d) Glutamine 40. Different initial conditions of H^N inversion recovery profiles (a,b) are indicated with the following symbols: (♦) non-selective, (▲) H^N-H^{water}-selective, (■) H^N-H^C-selective, and (●) H^N-selective. In the case of the H^α inversion recovery profiles (c,d) different initial conditions are indicated with the following symbols: (♦) non-selective, (►) H^C-selective, (■) H^N-H^C-selective, and (●) H^α-selective. The pH of the sample, and temperature at which experiments were conducted are specified in each graph.

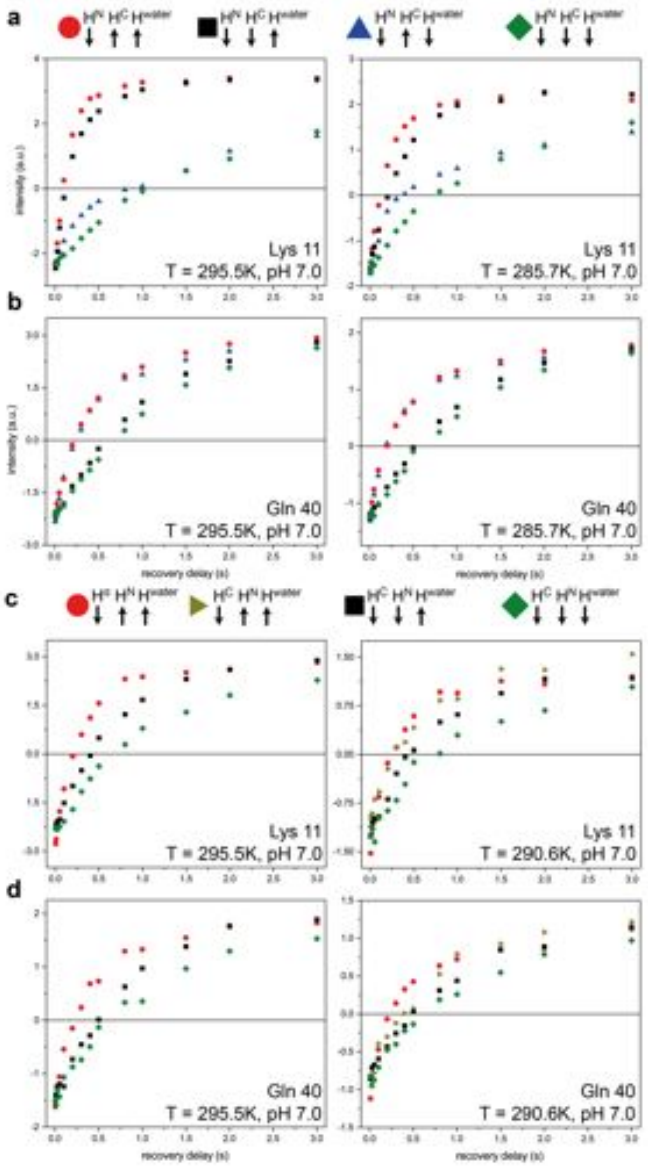


Figure S7. For the rough estimation of the effective T_1, the data obtained from the inversion recovery curves were fitted by single exponential function. H^N and H^α effective T_1 of α-synuclein. a) Different initial conditions are indicated with the following symbols: (♦) non-selective, (▲) H^N-H^{water}-selective, (■) H^N-H^C-selective, and (●) H^N-selective. b) In the case of the H^α effective T_1, different initial conditions are indicated with the following symbols: (♦) non-selective, (►) H^C-selective, (■) H^N-H^C-selective, and (●) H^α-selective. Sample conditions are indicated in the graphs.

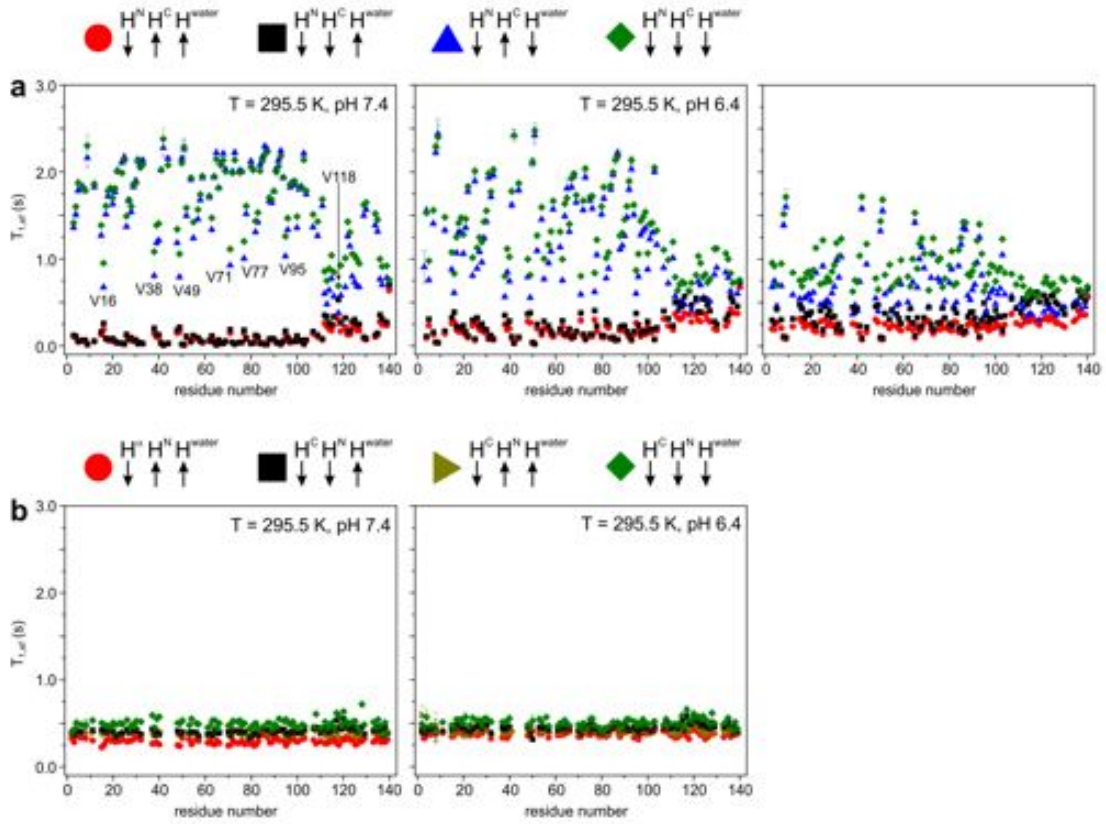
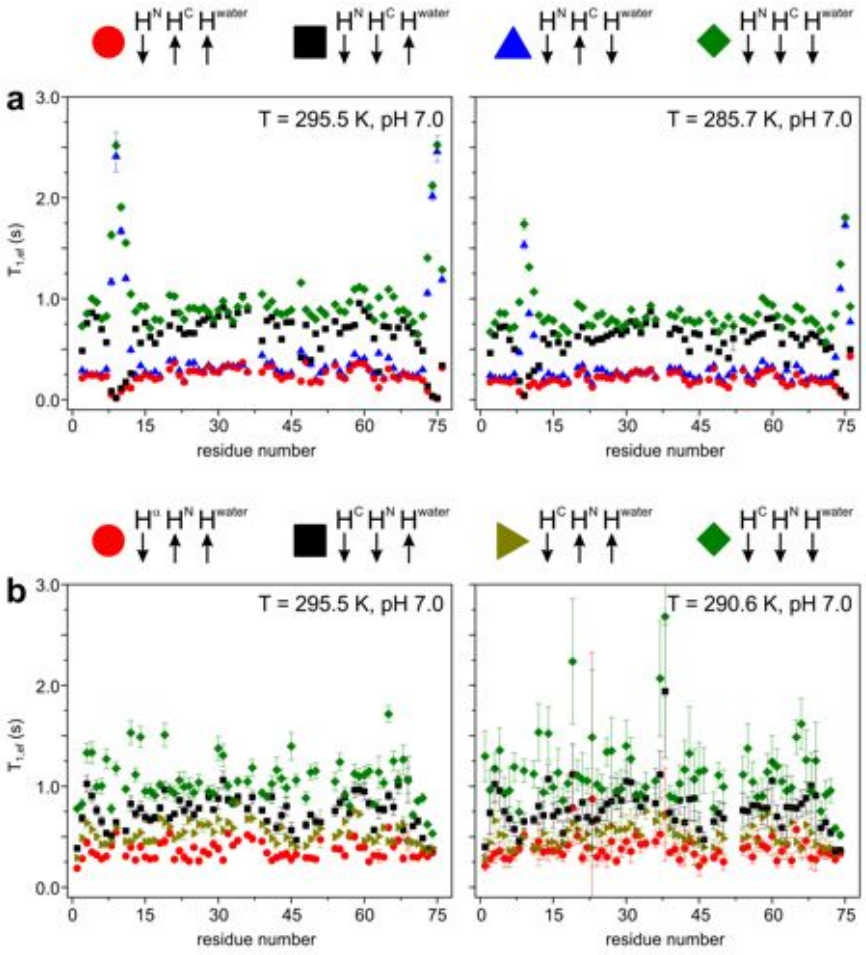


Figure S8. For the rough estimation of the effective T_1, the data obtained from the inversion recovery curves were fitted by single exponential function. H^N and H^α effective T_1 of ubiquitin. a) Different initial conditions are indicated with the following symbols: (♦) non-selective, (▲) H^N-H^{water}-selective, (■) H^N-H^C-selective, and (●) H^N-selective. b) In the case of the H^α effective T_1, different initial conditions are indicated with the following symbols: (♦) non-selective, (►) H^C-selective, (■) H^N-H^C-selective, and (●) H^α-selective. Sample conditions are indicated in the graphs.



5.1.3

"CON-CON" assignment strategy for highly flexible intrinsically disordered proteins

Alessandro Piai[a], Tomáš Hošek[a], Leonardo Gonnelli[a], Anna Zawadska-Kazimierczuk[b], Wiktor Koźmiński[b], Bernhard Brutscher[c], Wolfgang Bermel[d], Roberta Pierattelli[a], Isabella C. Felli[a]

[a] *CERM and Department of Chemistry Ugo Schiff, University of Florence, via Luigi Sacconi 6, Sesto Fiorentino, 50019, Italy*

[b] *Faculty of Chemistry, Biological and Chemical Research Centre, University of Warsaw, Żwirki I Wiguri 101, Warsaw, 02-089, Poland*

[c] *Institut de Biologie Structurale, Université Grenoble 1, CNRS, CEA Rue Jules Horowitz 41, Grenoble Cedex 1, 38027, France*

[d] *Bruker BioSpin AG, Industriestrasse 26, Fällanden, 8117, Switzerland*

Published in **2014** in *J. Biomol. NMR*, 60(4), 209-218

J Biomol NMR (2014) 60:209–218
DOI 10.1007/s10858-014-9867-6

ARTICLE

"CON-CON" assignment strategy for highly flexible intrinsically disordered proteins

Alessandro Piai · Tomáš Hošek · Leonardo Gonnelli · Anna Zawadzka-Kazimierczuk · Wiktor Koźmiński · Bernhard Brutscher · Wolfgang Bermel · Roberta Pierattelli · Isabella C. Felli

Received: 27 August 2014 / Accepted: 10 October 2014 / Published online: 19 October 2014

Abstract Intrinsically disordered proteins (IDPs) are a class of highly flexible proteins whose characterization by NMR spectroscopy is complicated by severe spectral overlaps. The development of experiments designed to facilitate the sequence-specific assignment procedure is thus very important to improve the tools for the characterization of IDPs and thus to be able to focus on IDPs of increasing size and complexity. Here, we present and describe the implementation of a set of novel ^{1}H-detected 5D experiments, (HACA)CON(CACO)NCO(CA)HA, BT-(H)NCO(CAN)CONNH and BT-HN(COCAN)CONNH, optimized for the study of highly flexible IDPs that exploit the best resolved correlations, those involving the carbonyl and nitrogen nuclei of neighboring amino acids, to achieve sequence-specific resonance assignment. Together with the analogous recently proposed pulse schemes based on ^{13}C detection, they form a complete set of experiments for sequence-specific assignment of highly flexible IDPs. Depending on the particular sample conditions (concentration, lifetime, pH, temperature, etc.), these experiments present certain advantages and disadvantages that will be discussed. Needless to say, that the availability of a variety of complementary experiments will be important for accurate determination of resonance frequencies in complex IDPs.

Keywords Intrinsically disordered proteins · ^{13}C detection · $^{1}H^{N}$ detection · $^{1}H^{\alpha}$ detection · NUS · Multidimensional NMR experiment · BEST-TROSY · Backbone assignment

Introduction

The role of intrinsic disorder is essential in many cellular mechanisms (Tompa 2009, 2012; Uversky 2013a, b), which range from gene regulation to signaling processes. Intrinsically disordered proteins (IDPs) or intrinsically disordered protein regions (IDPRs) play key roles in many of these cellular events (Uversky and Dunker 2013), thanks to their local flexibility and conformational freedom, which allow them to modulate such processes. These evidences are pushing the scientific community towards a reconsideration of the structure/function paradigm, taking into account that many disordered proteins are able to carry out specific functions despite the absence of secondary and tertiary structures (Wright and Dyson 1999; Uversky et al. 2000; Dunker et al. 2001).

With X-ray crystallography being incapable to describe conformational disorder, nuclear magnetic resonance (NMR) spectroscopy becomes the most qualified tool of

Electronic supplementary material The online version of this article (doi:10.1007/s10858-014-9867-6) contains supplementary material, which is available to authorized users.

A. Piai · T. Hošek · L. Gonnelli · R. Pierattelli (✉) · I. C. Felli (✉)
CERM and Department of Chemistry, University of Florence, Via Luigi Sacconi 6, 50019 Sesto Fiorentino, Florence, Italy
e-mail: felli@cerm.unifi.it

A. Zawadzka-Kazimierczuk · W. Koźmiński
Faculty of Chemistry, Biological and Chemical Research Centre, University of Warsaw, Żwirki i Wigury 101, 02-089 Warsaw, Poland

B. Brutscher
Institut de Biologie Structurale, Université Grenoble 1, CNRS, CEA, Rue Jules Horowitz 41, 38027 Grenoble Cedex 1, France

W. Bermel
Bruker BioSpin GmbH, Silberstreifen, 76287 Rheinstetten, Germany

investigation, thanks also to its unique ability to provide information at atomic resolution (Konrat 2014; Felli and Pierattelli 2014; Nováček et al. 2014). However, in respect to what is found for well-folded proteins, the peculiar properties of IDPs introduce additional challenges that need to be overcome to prevent drastic reduction in the quality of the spectra. First, the conformational freedom experienced by IDPs leads to severe collapse of chemical shifts, since their values are mostly determined by the amino acid chemical structure and by the primary sequence while contributions deriving from the three-dimensional structure are missing. Secondly, the high rates of solvent exchange processes endured by amide protons, that in absence of a stable fold are largely solvent exposed, can dramatically compromise the information content of multidimensional spectra based on $^{1}H^{N}$ detection, since at physiological temperature and pH many signals may be broadened even beyond detection. Finally, IDPs are often characterized by repetitive amino acid sequences in which prolines are usually abundant, thus complicating the resonance assignment procedure. To overcome these limitations as much as possible, several IDP-dedicated NMR experiments have been recently proposed (Panchal et al. 2001; Hiller et al. 2007; Mäntylahti et al. 2010; Motáčkova et al. 2010; Narayanan et al. 2010; Mäntylahti et al. 2011; Nováček et al. 2011; Favier and Brutscher 2011; Bermel et al. 2012; Zawadzka-Kazimierczuk et al. 2012b; Pantoja-Uceda and Santoro 2013a, b; Nováček et al. 2013; Bermel et al. 2013b; Kazimierczuk et al. 2013; Solyom et al. 2013; Pantoja-Uceda and Santoro 2014); however, NMR experiments able to deal with all these complications still need to be fully developed.

To cope with the low chemical shift dispersion of IDPs/IDPRs, high spectral resolution is mandatory. Therefore, increasing the number of dimensions of NMR experiments provides a unique tool to enhance the resolution, provided non-uniform sampling (NUS) (Kazimierczuk et al. 2010a, 2012) and, when possible, longitudinal relaxation enhancement (LRE) (Pervushin et al. 2002; Schanda and Brutscher 2005; Schanda et al. 2005; Deschamps and Campbell 2006; Felli and Brutscher 2009; Solyom et al. 2013; Gil et al. 2013) are used to achieve high resolution also in indirect dimensions without exponentially increasing the duration of NMR experiments. In fact, since the number of hypercomplex points correlates with the number of cross-peaks, by adding another dimension the required experimental time is usually doubled. The exploitation of the heteronuclei and in particular the chemical shift labeling of ^{15}N and ^{13}C frequencies provide great benefits to reduce spectral crowding (Ikura et al. 1991; Sattler et al. 1999; Dyson and Wright 2004; Mittag and Forman-Kay 2007). In our previous paper (Bermel et al. 2013b), we proposed three novel 5D ^{13}C direct-detected NMR experiments in which two pairs of ^{15}N and ^{13}C frequencies are correlated in order to provide robust information for fast and reliable backbone assignment. ^{13}C detection was chosen to exploit at best the greater chemical shift dispersion and the reduced sensitivity to exchange broadening of carbons in respect to protons (Csizmok et al. 2008; Pérez et al. 2009; Hsu et al. 2009; Knoblich et al. 2009; O'Hare et al. 2009; Skora et al. 2010; Felli et al. 2013). Here, we extend the same experimental strategy also to proton detection, combining the advantages given by heteronuclei in terms of spectral resolution with the direct acquisition of protons, which are characterized by about a 4-times larger gyromagnetic ratio resulting in higher intrinsic sensitivity. The new experiments should not be considered a replacement of the ^{13}C detected ones, but valid alternatives that can be adopted when appropriate, as discussed later. Indeed, we strongly believe that ^{1}H and ^{13}C detection provide complementary tools that allow to extend the size and complexity of IDPs/IDPRs that can be investigated by NMR. In fact, the chemical shift assignment is a mandatory step towards the structural and dynamic characterization of proteins, as well as of their interactions. The usefulness of the new experiments is demonstrated on a paradigmatic IDP, human α-synuclein (14,460 Da).

Materials and methods

NMR samples

A sample of 0.9 mM uniformly ^{13}C, ^{15}N labeled human α-synuclein in 20 mM phosphate buffer at pH 6.5 was prepared as previously described (Huang et al. 2005). EDTA and NaCl were added to reach the final concentration of 0.5 and 200 mM, respectively, and 10 % D_2O was added for the lock (Sample *A*). An identical sample, but in fully deuterated solvent (99.99 % D_2O), was also prepared (Sample *B*). The H^{N} and H^{α}-based experiments were acquired on sample *A* and *B*, respectively, both in 3 mm NMR sample tubes to reduce the detrimental effects of high salt concentration. The temperature was set to 285.5 K for Sample *A*, to reduce the rate of chemical exchange of amide protons, and to 310.0 K for Sample *B*, to work as close as possible to physiological conditions.

NMR data acquisition

All the NMR experiments were performed at 22.3 T on a Bruker Avance III spectrometer operating at 950.20 MHz ^{1}H, 238.93 MHz ^{13}C and 96.28 MHz ^{15}N frequencies, equipped with a cryogenically cooled probehead. Parameters specific to each experiment are reported in the captions of the figures describing the pulse sequences, all reported in

the Supplementary Material. E-BURP2 (or time reversed E-BURP2) (Geen and Freeman 1991) and REBURP (Geen and Freeman 1991) shapes of durations of 1,200 and 1,180 μs, respectively, were employed for ^{1}H band-selective π/2 and π flip angle pulses; BIP-750-50-20 pulse shapes (Smith et al. 2001) of duration of 200 μs were used for broadband ^{1}H inversion. For ^{13}C band-selective π/2 and π flip angle pulses G4 (or time reversed G4) (Emsley and Bodenhausen 1990) and Q3 shapes (Emsley and Bodenhausen 1992) of durations of 260 and 161 μs, respectively, were used, except for the π pulses that should be band-selective on the C^{α} region (Q3, 667 μs) and for the adiabatic π pulse to invert both C′ and C^{α} (smoothed Chirp 500 μs, 20 % smoothing, 80 kHz sweep width, 11.3 kHz RF field strength) (Böhlen and Bodenhausen 1993). The ^{13}C band selective pulses on C^{α} and C′ were applied at the center of each region, respectively. For the $^{1}H^{\alpha}$ experiments, decoupling of ^{1}H and ^{13}C was achieved with DIPSI-2 (Shaka et al. 1988; Cavanagh and Rance 1992) (3.1 kHz) and GARP-4 (Shaka et al. 1985) (4.5 kHz) sequences, respectively. All gradients employed had a smoothed square shape. The parameters used for the acquisition of the 3D and 5D experiments are reported, respectively, in Table 1 and 2. Each experiment was acquired with the States or Echo-Antiecho method applied in the indirect dimensions to achieve quadrature detection. All the experiments were performed using on-grid non-uniform sampling (NUS). The on-grid "Poisson disk" sampling scheme (Kazimierczuk et al. 2008) was chosen to generate the time schedules with the *RSPack* program. The distribution was relaxation-optimized, i.e. the density of points was decaying according to the Gaussian distribution exp $(-t^2/\sigma^2)$, with $\sigma = 0.5$. All the spectra were acquired using *Bruker TopSpin 3.1* software.

NMR data processing and analysis

The experimental data were converted with *nmrPipe* (Delaglio et al. 1995). 3D data were processed using the multidimensional Fourier transform (MFT) algorithm implemented in the *ToASTD* program (Kazimierczuk et al. 2006). 5D data were processed using the Sparse MFT (SMFT) algorithm implemented in the *reduced* program (Kazimierczuk et al. 2009, 2010b). Both programs are available at http://nmr.cent3.uw.edu.pl/software. Finally, *Sparky* (Goddard and Kneller 2000) was used to analyze the spectra.

Handling and inspection of 5D NMR spectra

The SMFT algorithm (Kazimierczuk et al. 2009) was employed to process the 5D data sets. This method enables to process data with very high resolution in all dimensions by storing only the informative parts of multidimensional

Table 1 Experimental parameters used for the 3D experiments (providing the basis spectra)

	Spectral widths and maximal evolution times			No. of scans	Inter-scan delays (s)	No. of complex points (aq)	No. of hypercomplex points	Duration of the experiment	Relative data points density (%)
3D (HACACO)NCO (CA)HA	3.5 kHz (^{15}N) 35.7 ms	2.0 kHz (^{13}C′) 35.0 ms	13.3 kHz ($^{1}H^{\alpha}$)	4	1.0	1,024	1,450	8 h	20.3
5D BEST TROSY HNCO	1.8 kHz (^{13}C′) 41.7 ms	2.6 kHz (^{15}N) 40.4 ms	13.3 kHz ($^{1}H^{N}$)	4	0.2	1,024	1,575	3 h	20.0

Table 2 Experimental parameters used for the 5D experiments

	Spectral widths and maximal evolution times					No. of scans	Inter-scan delays (s)	No. of complex points (aq)	No. of hyper-complex points	Duration of the experiment	Relative data points density (%)
	Cross-sections		Dimensions shared with the basis spectrum								
5D (HACA) CON(CACO) NCO(CA)HA	2.0 kHz (^{13}C′) 35.0 ms	3.5 kHz (^{15}N) 35.7 ms	3.5 kHz (^{15}N) 35.7 ms	2.0 kHz (^{13}C′) 35.0 ms	13.3 kHz ($^{1}H^{\alpha}$)	8	1.0	1,024	1,150	2 days, 7 h	0.0015
5D BEST TROSY (H)NCO (CAN)CONNH	2.6 kHz (^{15}N) 40.4 ms	1.8 kHz (^{13}C′) 41.7 ms	1.8 kHz (^{13}C′) 41.7 ms	2.6 kHz (^{15}N) 40.4 ms	13.3 kHz ($^{1}H^{N}$)	4	0.2	1,024	1,200	11 h	0.0019
5D BEST TROSY HN (COCAN) CONNH	2.0 kHz ($^{1}H^{N}$) 40.0 ms	2.6 kHz (^{15}N) 40.4 ms	1.8 kHz (^{13}C′) 41.7 ms	2.6 kHz (^{15}N) 40.4 ms	13.3 kHz ($^{1}H^{N}$)	4	0.2	1,024	1,200	11 h	0.0018

spectra. Indeed processing NMR spectra with high resolution in all dimensions would provide prohibitively large data files for spectra with more than three dimensions. However, in most cases the number of expected cross-peaks in the spectra does not increase passing from a 3D to a higher dimensional experiment. Therefore, these high dimensional spaces are essentially "empty" with interesting information confined in very narrow regions, those where cross-peaks are observed in the lower-dimensionality 3D spectra. One interesting feature of SMFT is its unique ability to simplify the analysis of multidimensional spectra, by considering that a 5D spectrum can be decomposed into a 3D spectrum in which each detected correlation peak is associated to an additional 2D spectrum, contributing the remaining two chemical shifts. Actually, the 3D spectrum (to which we refer as "basis spectrum") can be conveniently obtained from a simpler (and thus also more sensitive) 3D experiment correlating three (out of the five) resonances. The two remaining frequencies, which are the new information-content of the 5D spectrum, are retrieved in a series of 2D spectra (to which we refer as "cross-sections"), which are the only spectra computed from the 5D data sets. Therefore, the analysis of a 5D spectrum is simplified to the inspection of a series of 2D cross-sections, to which the three "basic" frequencies, retrieved in the corresponding 3D spectrum, are joined.

Results and discussion

To acquire high dimensional NMR experiments (4D, 5D, etc.), the use of sparse sampling strategies is very convenient. In fact, data sampling at non-uniform intervals is not only an ingenious way to speed up the acquisition of multidimensional NMR experiments, but mostly a powerful method to significantly increase the achievable spectral resolution of the indirect dimensions, as demanded by IDPs. Therefore, all the spectra were acquired using NUS. An inconvenient feature of NUS data processing is the appearance of sampling artifacts in the spectrum. Sampling artifacts can be efficiently suppressed, for example by employing the signal separation algorithm (SSA) (Stanek et al. 2012). However, in the case of protein backbone assignment, with spectra characterized by a small number of correlation peaks, artifacts are of low intensity and can be safely ignored.

Using the experimental strategy outlined in the "Materials and methods" section, we recently built a set of 4D and 5D NMR experiments based on ^{13}C-direct detection, specifically tailored to facilitate the chemical shift assignment of IDPs (Bermel et al. 2012, 2013b). The high dimensional experiments were designed as expansion of the well-known 2D CON-IPAP and 3D CACON-IPAP experiments (used as basis spectra to process the 4D and 5D data sets, respectively). From our experience, the experiments in which two pairs of ^{15}N and ^{13}C′ frequencies belonging to two neighboring residues are correlated, in order to sequentially link the peaks retrieved in the 2D CON-IPAP spectrum, provide the best result in terms of rapidity and reliability of the sequential backbone assignment (Bermel et al. 2013b; Pantoja-Uceda and Santoro 2014). This assignment approach, to which we like to refer as "CON-CON strategy", takes advantage of the relatively good chemical shift dispersion of ^{15}N and ^{13}C′ frequencies (Bermel et al. 2013a) of IDPs, which significantly improve the resolution of the spectra together with the use of three or four indirect dimensions. The employment of ^{13}C-direct detection provides a series of additional benefits. First, it reduces the minimum number of dimensions needed to apply the CON-CON strategy to four, because one of the two ^{13}C′ dimensions is directly acquired. Secondly, it allows to detect correlation peaks for all residues, included prolines that are missing in amide ^{1}H-detected experiments. Finally, it allows to perform experiments at near physiological conditions (high temperature and pH), since carbon nuclei are not directly affected by chemical exchange with the solvent. The major drawback of ^{13}C′ direct-detected experiments is the ^{13}C intrinsic lower sensitivity, largely compensated by recent technological improvements, which however can still become a limiting factor for complete assignment of less-concentrated or short-lived protein samples. Herein we now propose a set of 5D NMR experiments that still exploit the CON-CON strategy, but make use of protons instead of carbons for detection.

In order to take advantage of the particular properties of aliphatic and amide ^{1}H, we developed two different pulse sequences in which either H^{α} or H^{N} nuclei are exploited for detection as well as for the starting source of magnetization. The new 5D experiments, named (HACA)CON(CACO)NCO(CA)HA, BT-(H)NCO(CAN)CONNH (and its variant BT-HN(COCAN)CONNH) will be discussed in the following in more detail. In all experiments, pairs of ^{15}N and ^{13}C′ resonances are collected in the indirect dimensions. Semi-constant time or real time chemical shift evolution is used to achieve highest possible spectral resolution in all dimensions, making transverse relaxation the only limitation for choosing the appropriate acquisition time.

For its similarity with carbon detection, alpha proton detection (Mäntylahti et al. 2011) was chosen in the design of the 5D (HACA)CON(CACO)NCO(CA)HA experiment. In fact, H^{α} are not influenced by chemical exchange with the solvent even at high temperature and pH, thus being very appropriate to investigate IDPs at near physiological conditions. Moreover, H^{α}-based experiments provide direct information on proline residues, giving the possibility to

obtain virtually complete assignment. The use of deuterated water as solvent (99.99 % D_2O) ensures to completely prevent missing signals under the residual water peak. It also provides further advantages that can enhance the spectrometer performance, like the possibility to work at receiver gain values where digitizer noise is not an issue and to avoid all the phenomena related to radiation damping which become an issue when working with cryoprobes and high salt sample conditions, often necessary when dealing with IDPs. A slight drawback is the increased viscosity of D_2O with respect to H_2O that results in increased transverse relaxation rates. This effect is, however, counterbalanced by the possibility to work at higher temperature, as there is no need to slow down solvent exchange processes.

The coherence transfer pathway and the information provided by the 5D (HACA)CON(CACO)NCO(CA)HA experiment are reported in Fig. 1, top panel. The flow of magnetization, which starts and ends on alpha protons, is established through ^{1}J scalar coupling transfer steps; in particular, the back-transfer step mediated by $^2J(N–C^\alpha)$ scalar coupling is suppressed. Therefore, the experiment correlates C'_{i-1}–N_i–N_{i+1}–C'_i–H_i^α nuclei, making the magnetization transfer pathway unidirectional. In this way, the sequence-specific backbone assignment procedure becomes a "walk" through a series of CON cross-peaks. The sequence was designed to detect only one peak in each cross-section, reducing at maximum the risk of peak overlaps. For SMFT processing, a 3D (HACACO)NCO(CA)HA experiment (Mäntylahti et al. 2010) has to be performed, as it provides N_{i+1}–C'_i–H_i^α frequency correlations (basis spectrum), that are used to extract from the 5D spectrum the additional C'_{i-1}–N_i frequencies. Hence, the "CON-CON strategy" is carried out sequence-specifically linking the pairs of C'_i and N_{i+1} resonances, retrieved in the cross-sections, to the ones present in the basis spectrum, as shown in Fig. 2. To facilitate the mapping of the linked resonances on the amino acidic sequence, the 5D (HACA)CON(CACO)NCO(CA)HA experiment has been designed in a way that cross peaks for glycine residues (at the *i* position) are of opposite sign with respect to all others, as illustrated in Fig. 3 where six cross-sections of the 5D spectrum are shown. It is interesting to note that, with the resolution obtained in the direct ^{1}H dimension, doublets due to $^3J(H^\alpha–H^\beta)$ scalar couplings are detected for alanine, valine, isoleucine and threonine residues. However, these line splittings do not significantly complicate the analysis of the spectra.

As an alternative to H^α detection, the 5D BT-(H)NCO(CAN)CONNH experiment was developed to exploit the same "CON-CON strategy", but using amide proton detection. The coherence transfer pathway and the information provided by this experiment are reported in Fig. 1, bottom panel. The experiment is a variant of the previously published 5D (H)NCO(NCA)CONH experiment (Zawadzka-Kazimierczuk et al. 2012b). However, in our new pulse sequence, the magnetization is transferred from and to neighboring amide protons by again only exploiting ^{1}J scalar coupling transfer steps, resulting in unambiguous N_{i+1}–C'_i–C'_{i-1}–N_i–H_i^N correlations. The experiment is implemented as a BEST-TROSY (BT) version, in order to benefit from the short selective longitudinal relaxation times of amide protons in IDPs (Solyom et al. 2013; Gil et al. 2013). The combination of BEST (Pervushin et al. 2002; Schanda 2009; Schanda et al. 2006) and TROSY (Pervushin et al. 1997), as described by Favier and Brutscher (Favier and Brutscher 2011), ensures highest

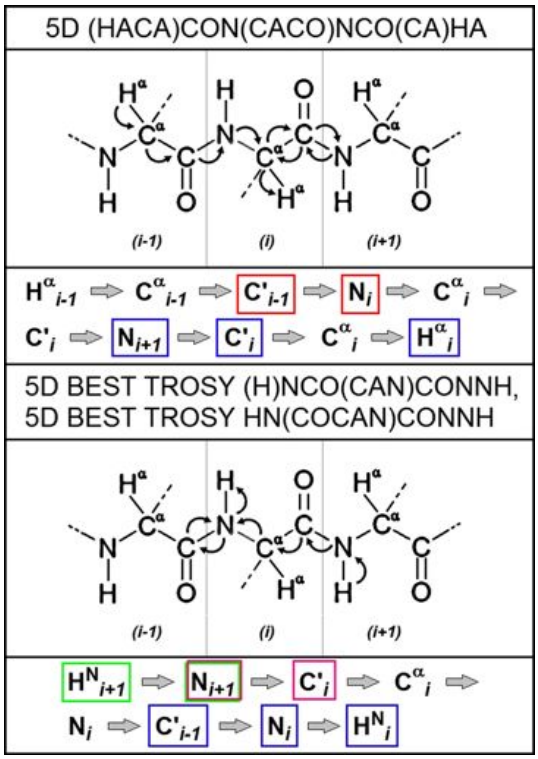


Fig. 1 The flow of the magnetization in the proposed experiments is schematically illustrated. The magnetization transfer pathway of the 5D (HACA)CON(CACO)NCO(CA)HA experiment is reported in the *top panel*, whereas that of the 5D BT-(H)NCO(CAN)CONNH and BT-HN(COCAN)CONNH experiments at the *bottom*. All *arrows* represent coherence transfer steps through ^{1}J scalar couplings. For each experiment, the correlations retrieved are also listed explicitly. Those embedded in colored boxes are the ones frequency-labeled within the experiments. *Blue rectangles* show the dimensions shared with the basis spectrum, whereas *red*, *magenta* and *green rectangles* indicate the dimensions of the 2D cross-sections of the 5D (HACA)CON(CACO)NCO(CA)HA, 5D BT-(H)NCO(CAN)CONNH and 5D BT-HN(COCAN)CONNH experiments, respectively. The two variants of the BEST TROSY experiments have in common four dimensions

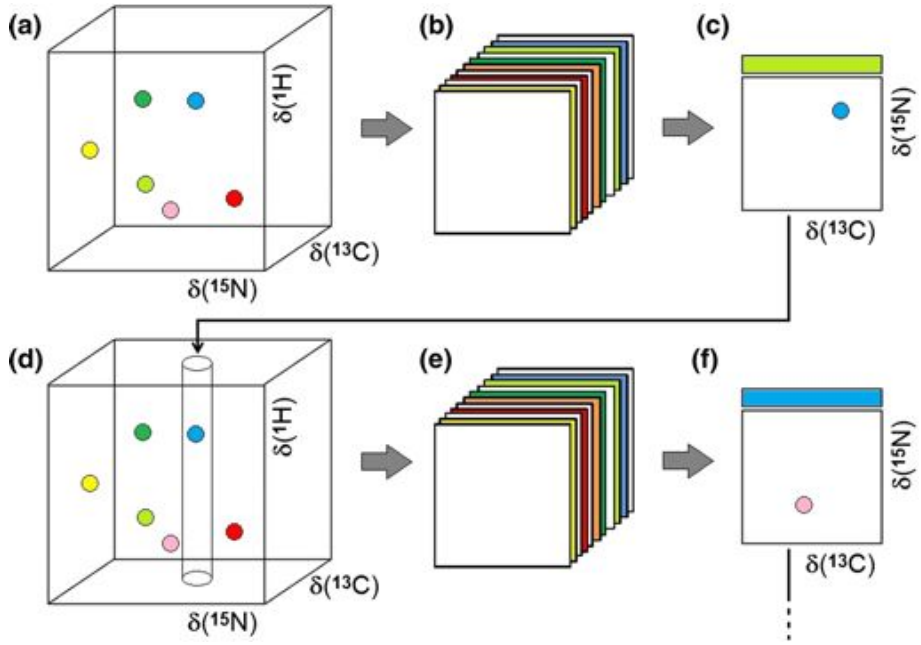


Fig. 2 Schematic illustration of the "CON-CON strategy". The sequence-specific assignment is performed through simultaneous exploitation of C′ and N resonances, linked between the 2D cross-sections of the 5D spectrum and the 3D basis spectrum. The same procedure can be used both for $^{1}H^{\alpha}$ and $^{1}H^{N}$ detected spectra, as described here schematically. First, a cross-peak (*light green* in the example) of the 3D basis spectrum is chosen (**a**). Thereafter, the related 2D cross-section of the 5D spectrum is inspected (**b**) to retrieve the frequencies of the CON peak immediately preceding or succeeding (for the $^{1}H^{\alpha}$ and $^{1}H^{N}$ spectrum, respectively) the one of the basis spectrum (**c**). Once the cross-peak (*blue* in the example) of the 3D basis spectrum with the same C′ and N frequencies is found (**d**), it becomes sequence-specifically linked to the one found in (**a**). Then, the procedure continues in (**e**) and (**f**) as explained in (**b**) and (**c**), forming chains of adjacent CON peaks in the primary sequence of the protein

experimental sensitivity and spectral resolution in very short overall acquisition times. As amide protons are affected by chemical exchange with the solvent, it is opportune to perform the experiment at lower temperature and pH in order to prevent extensive line broadening. In order to obtain the basis spectrum, correlating C'_{i-1}–N_i–H_i^N nuclei, an additional BT-HNCO (Solyom et al. 2013) spectrum is recorded. The frequency information, obtained from BT-HNCO, then allows to extract a series of 2D cross-sections from the 5D BT-(H)NCO(CAN)CONNH data set that adds the additional N_{i+1}–C'_i information (Fig. 4, panel a). Then, as described for the H^{α}-detected experiment, it is possible to establish sequential correlations linking the pairs of C'_i and N_{i+1} frequencies present in the 2D cross-sections of the 5D spectrum to those embedded in the 3D basis spectrum (Fig. 2). The only difference with the H^{α}-detected experiment is that the sequential walk now proceeds from the N- to the C-terminus, instead of the inverse. Because of the higher intrinsic sensitivity and shorter overall experimental time requirements, the 5D BT-(H)NCO(CAN)CONNH experiments is especially appealing for fast-degrading or low-concentrated IDP samples.

Finally, the above described "CON-CON strategy" can be extended to include the amide ^{1}H, thus becoming a "HNCO-HNCO strategy". For this purpose, the H^N_{i+1} needs to be added to the frequency information contained in the 5D BT-(H)NCO(CAN)CONNH spectrum. In principle, this can be achieved by recording a 6D BT-HNCO(CAN)CONNH correlation spectrum, adding an amide ^{1}H chemical shift evolution period at the beginning of the pulse sequence (Zawadzka-Kazimierczuk et al. 2012b). Here we have chosen an alternative approach that consists in recording a second 5D experiment, BT-HN(COCAN)CONNH, where chemical shift evolution of one of the carbonyls is replaced by that of the remaining amide proton, thus correlating H^N_{i+1}–N_{i+1}–C'_{i-1}–N_i–H_i^N nuclei (Fig. 1, bottom panel). The same 3D BT-HNCO spectrum can be used to compute the 2D cross-sections for this additional 5D experiment. Recording of the two 5D data sets with a different set of frequency-labeled nuclei is equivalent to recording a complete 6D data set as long as there is no degenerate N_{i+1}–C'_{i-1}–N_i–H_i^N frequency correlation. This strategy is well-known as projection spectroscopy (Kupce and Freeman 2003). The "HNCO-HNCO strategy" is likely to provide a high level of unambiguous sequential

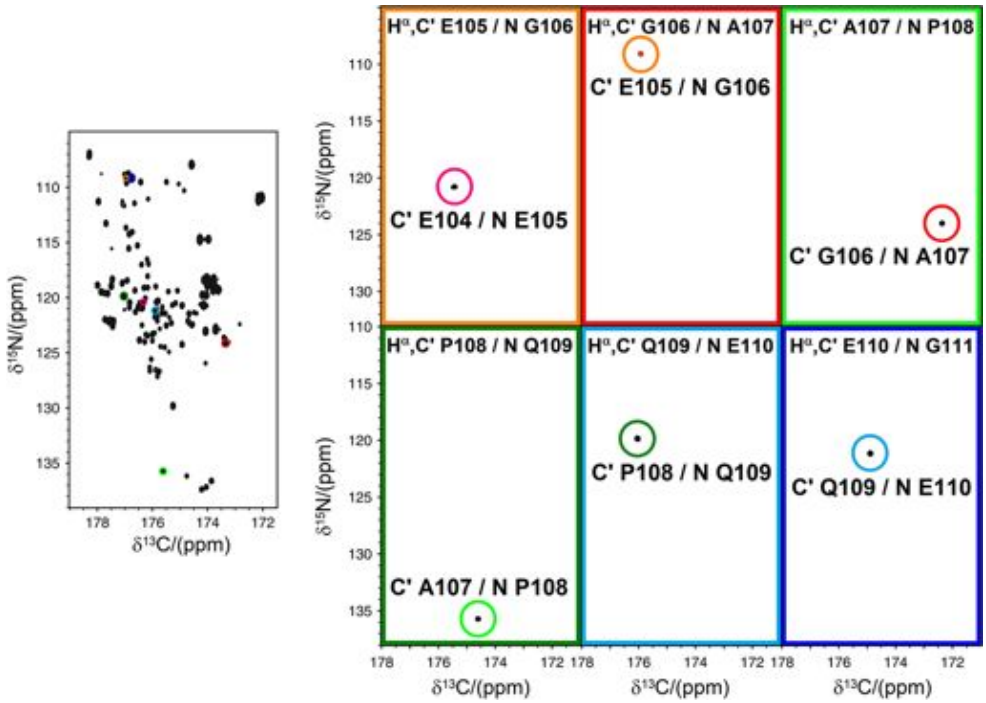


Fig. 3 Six 2D cross-sections of the 5D (HACA)CON(CACO)NCO (CA)HA spectrum are shown (*right side*), extracted at the N_{i+1}–C'_i–H^{α}_i frequencies reported on the top of each panel and schematically indicated by circles on the 2D C′–N projection of the 3D basis spectrum. In each cross-section, the observed cross-peak allows to identify the frequencies of the previous CON cross-peak (N_i–C'_{i-1}). *Black* peaks are positive, *red* negative. Inversion of peak sign is expected when a glycine is frequency-labeled. Therefore, negative peaks allow a straightforward identification of glycines. The sequential specific assignment is performed connecting the CON peak of a given cross-sections to the one of the basis spectrum with identical C′ and N frequencies, as explained in Fig. 2 and illustrated here with the use of colors, so that each color identifies a specific pair of N_{i+1}–C'_i frequencies. The simultaneous exploitation of carbonyl carbon and nitrogen makes the sequential assignment extremely reliable. As clearly visible, the use of five dimensions provide extremely-resolved cross-sections, minimizing the chance of possible overlaps

assignment, even for large IDPs with repetitive primary sequences, since it is very rare even for IDPs that two or more peaks retrieved in the 3D HNCO spectrum are completely overlapping. Analogously, a similar strategy could also be considered for the H^{α}-detected experiment although the added value is less pronounced due to the lower chemical shift dispersion of H^{α}.

Finally, it's worth mentioning that sequence-specific assignment of resonances based on the proposed experiments can be performed automatically, for example by using the *TSAR* program (Zawadzka-Kazimierczuk et al. 2012a), available at http://nmr.cent3.uw.edu.pl/software. However, the programs for automatic assignment usually require at least one experiment providing sequential connectivities and at least one experiment providing information on amino-acid type. Therefore automatic assignment procedures were not attempted in the present work, since all proposed experiments belong to the first category.

Conclusions

We have presented new high-dimensional (5D) experiments exploiting C'_i–N_{i+1} pairs for sequence-specific assignment, based on either H^{α} or H^{N} detection. These experiments provide new complementary tools to the previously described ^{13}C-detected experiments for highly efficient backbone resonance assignment of flexible IDPs. The "CON-CON strategy", including different experimental variants based on ^{13}C, $^{1}H^{\alpha}$ and $^{1}H^{N}$ detection, and its extension to the "HNCO-HNCO strategy", in case of $^{1}H^{N}$ detection, allow to establish almost unambiguous sequential correlations along the protein backbone. While the H^{α}–detected experiment is especially useful for IDP samples studied under near physiological conditions and/or characterized by proline-rich repetitive amino acid sequences, the H^{N}-detected experiments are expected to be favorable for low concentrated or short-lived IDP samples that can be studied under conditions where solvent exchange is sufficiently slowed down.

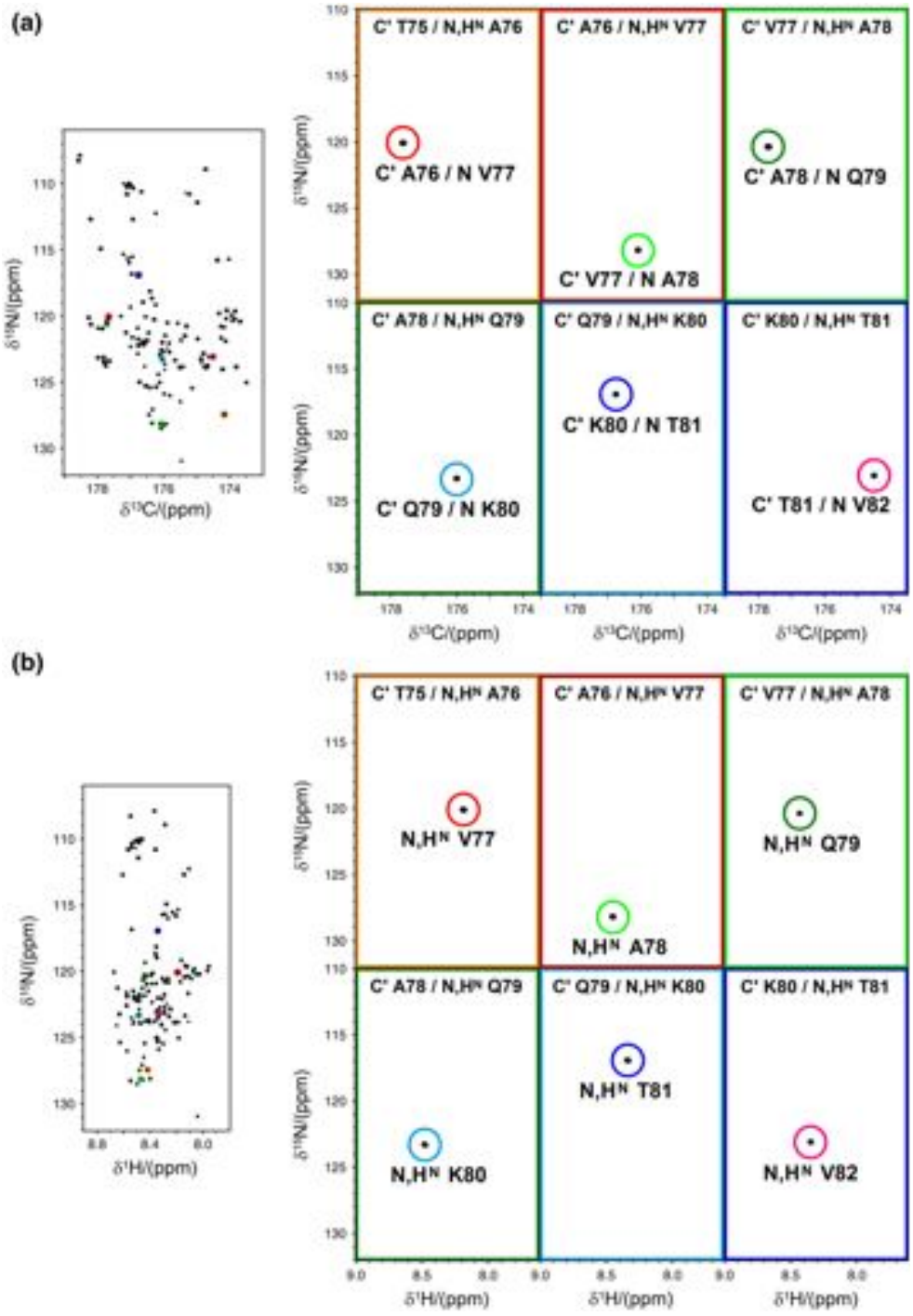


Fig. 4 Six 2D cross-sections of the 5D BT-(H)NCO(CAN)CONNH and 5D BT-HN(COCAN)CONNH spectra, extracted at C'_{i-1}–N_i–H^N_i frequencies, are shown respectively in (**a**) and (**b**). The color code and all the labels have the same meaning as described in the caption of Fig. 3. In each cross-section, the observed cross-peak allows to identify (**a**) the succeeding CON (C'_i–N_{i+1}) and (**b**) the succeeding HN (N_{i+1}–H^N_{i+1}). For this reason, on the left side the related C′–N (**a**) and H^N–N (**b**) projection of the common 3D HNCO basis experiment are reported. Since the cross-section of the two spectra share one nitrogen dimension, they can be exploited together to provide almost unambiguous assignment in the so-called "HNCO-HNCO strategy"

Acknowledgments This work has been supported in part by the European Commission Projects IDPbyNMR (Contract No. 264257), BioNMR (Contract No. 261863) and INSTRUCT (Contract No. 211252).

References

Bermel W, Bertini I, Gonnelli L, Felli IC, Koźmiński W, Piai A, Pierattelli R, Stanek J (2012) Speeding up sequence specific assignment of IDPs. J Biomol NMR 53:293–301

Bermel W, Bruix M, Felli IC, Kumar VMV, Pierattelli R, Serrano S (2013a) Improving the chemical shift dispersion of multidimensional NMR spectra of intrinsically disordered proteins. J Biomol NMR 55:231–237

Bermel W, Felli IC, Gonnelli L, Koźmiński W, Piai A, Pierattelli R, Zawadzka-Kazimierczuk A (2013b) High-dimensionality ^{13}C direct-detected NMR experiments for the automatic assignment of intrinsically disordered proteins. J Biomol NMR 57:353–361

Böhlen J-M, Bodenhausen G (1993) Experimental aspects of chirp NMR spectroscopy. J Magn Reson Ser A 102:293–301

Cavanagh J, Rance M (1992) Suppression of cross-relaxation effects in TOCSY spectra via a modified DIPSI-2 mixing sequence. J Magn Reson 96:670–678

Csizmok V, Felli IC, Tompa P, Banci L, Bertini I (2008) Structural and dynamic characterization of intrinsically disordered human securin by NMR. J Am Chem Soc 130:16873–16879

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 6:277–293

Deschamps M, Campbell ID (2006) Cooling overall spin temperature: protein NMR experiments optimized for longitudinal relaxation effects. J Magn Reson 178:206–211

Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Ratliff CM, Hipps KW, Ausio J, Nissen MS, Reeves R, Kang C, Kissinger CR, Bailey RW, Griswold MD, Chiu W, Garner EC (2001) Intrinsically disordered protein. J Mol Graph Model 19:26–59

Dyson HJ, Wright PE (2004) Unfolded proteins and protein folding studied by NMR. Chem Rev 104:3607–3622

Emsley L, Bodenhausen G (1990) Gaussian pulse cascades: new analytical functions for rectangular selective inversion and in-phase excitation in NMR. Chem Phys Lett 165:469–476

Emsley L, Bodenhausen G (1992) Optimization of shaped selective pulses for NMR using a quaternion description of their overall propagators. J Magn Reson 97:135–148

Favier A, Brutscher B (2011) Recovering lost magnetization: polarization enhancement in biomolecular NMR. J Biomol NMR 49:9–15

Felli IC, Brutscher B (2009) Recent advancements in solution NMR: fast methods and heteronuclear direct detection. ChemPhysChem 10:1356–1368

Felli IC, Pierattelli R (2014) Novel methods based on ^{13}C detection to study intrinsically disordered proteins. J Magn Reson 241:115–125

Felli IC, Piai A, Pierattelli R (2013) Recent advances in solution NMR studies: ^{13}C direct detection for biomolecular NMR applications. Ann Rep NMR Spectroscop 80:359–418

Geen H, Freeman R (1991) Band-selective radiofrequency pulses. J Magn Reson 93:93–141

Gil S, Hošek T, Solyom Z, Kümmerle R, Brutscher B, Pierattelli R, Felli IC (2013) NMR studies of intrinsically disordered proteins near physiological conditions. Angew Chem Int Ed 52:11808–11812

Goddard TD, Kneller DG (2000) SPARKY 3. University of California, San Francisco

Hiller S, Wasmer C, Wider G, Wüthrich K (2007) Sequence-specific resonance assignment of soluble nonglobular proteins by 7D APSY-NMR spectroscopy. J Am Chem Soc 129:10823–10828

Hsu ST, Bertoncini CW, Dobson CM (2009) Use of protonless NMR spectroscopy to alleviate the loss of information resulting from exchange-broadening. J Am Chem Soc 131:7222–7223

Huang C, Ren G, Zhou H, Wang C (2005) A new method for purification of recombinant human alpha-synuclein in *Escherichia coli*. Protein Expr Purif 42:173–177

Ikura M, Spera S, Barbato G, Kay LE, Krinks M, Bax A (1991) Secondary structure and side-chain ^{1}H and ^{13}C resonance assignments of calmodulin in solution by heteronuclear multidimensional NMR spectroscopy. Biochemistry 30:9216–9228

Kazimierczuk K, Zawadzka A, Koźmiński W, Zhukov I (2006) Random sampling of evolution time space and Fourier transform processing. J Biomol NMR 36:157–168

Kazimierczuk K, Zawadzka A, Koźmiński W (2008) Optimization of random time domain sampling in multidimensional NMR. J Magn Reson 192:123–130

Kazimierczuk K, Zawadzka A, Koźmiński W (2009) Narrow peaks and high dimensionalities: exploiting the advantages of random sampling. J Magn Reson 197:219–228

Kazimierczuk K, Stanek J, Zawadzka-Kazimierczuk A, Koźmiński W (2010a) Random sampling in multidimensional NMR spectroscopy. Prog NMR Spectrosc 57:420–434

Kazimierczuk K, Zawadzka-Kazimierczuk A, Koźmiński W (2010b) Non-uniform frequency domain for optimal exploitation of non-uniform sampling. J Magn Reson 205:286–292

Kazimierczuk K, Misiak M, Stanek J, Zawadzka-Kazimierczuk A, Koźmiński W (2012) Generalized Fourier transform for non-uniform sampled data. Top Curr Chem 316:79–124

Kazimierczuk K, Stanek J, Zawadzka-Kazimierczuk A, Koźmiński W (2013) High-dimensional NMR spectra for structural studies of biomolecules. ChemPhysChem 14:3015–3025

Knoblich K, Whittaker S, Ludwig C, Michiels P, Jiang T, Schaffhausen B, Günther U (2009) Backbone assignment of the N-terminal polyomavirus large T antigen. Biomol NMR Assign 3:119–123

Konrat R (2014) NMR contributions to structural dynamics studies of intrinsically disordered proteins. J Magn Reson 241:74–85

Kupce E, Freeman R (2003) Projection-reconstruction of three-dimensional NMR spectra. J Am Chem Soc 125:13958–13959

Mäntylahti S, Aitio O, Hellman M, Permi P (2010) HA-detected experiments for the backbone assignment of intrinsically disordered proteins. J Biomol NMR 47:171–181

Mäntylahti S, Hellman M, Permi P (2011) Extension of the HA-detection based approach: (HCA)CON(CA)H and (HCA)NCO(CA)H experiments for the main-chain assignment of intrinsically disordered proteins. J Biomol NMR 49:99–109

Mittag T, Forman-Kay J (2007) Atomic-level characterization of disordered protein ensembles. Curr Opin Struct Biol 17:3–14

Motáčkova V, Nováček J, Zawadzka-Kazimierczuk A, Kazimierczuk K, Žídek L, Sanderová H, Krásný L, Koźmiński W, Sklenář V (2010) Strategy for complete NMR assignment of disordered proteins with highly repetitive sequences based on resolution-enhanced 5D experiments. J Biomol NMR 48:169–177

Narayanan RL, Dürr HN, Bilbow S, Biernat J, Mendelkow E, Zweckstetter M (2010) Automatic assignment of the intrinsically disordered protein Tau with 441-residues. J Am Chem Soc 132:11906–11907

Nováček J, Zawadzka-Kazimierczuk A, Papoušková V, Žídek L, Sanderová H, Krásný L, Koźmiński W, Sklenář V (2011) 5D 13C-detected experiments for backbone assignment of unstructured proteins with a very low signal dispersion. J Biomol NMR 50:1–11

Nováček J, Janda L, Dopitová R, Žídek L, Sklenář V (2013) Efficient protocol for backbone and side-chain assignments of large, intrinsically disordered proteins: transient secondary structure analysis of 49.2 kDa microtubule associated protein 2c. J Biomol NMR 56:291–301

Nováček J, Žídek L, Sklenář V (2014) Toward optimal-resolution NMR of intrinsically disordered proteins. J Magn Reson 241:41–52

O'Hare B, Benesi AJ, Showalter SA (2009) Incorporating 1H chemical shift determination into 13C-direct detected spectroscopy of intrinsically disordered proteins in solution. J Magn Reson 200:354–358

Panchal SC, Bhavesh NS, Hosur RV (2001) Improved 3D triple resonance experiments, HNN and HN(C)N, for H^N and ^{15}N sequential correlations (^{13}C, ^{15}N) labeled proteins: application to unfolded proteins. J Biomol NMR 20:135–147

Pantoja-Uceda D, Santoro J (2013a) A suite of amino acid residue type classification pulse sequences for ^{13}C-detected NMR of proteins. J Magn Reson 234:190–196

Pantoja-Uceda D, Santoro J (2013b) Direct correlation of consecutive C′–N groups in proteins: a method for the assignment of intrinsically disordered proteins. J Biomol NMR 57:57–63

Pantoja-Uceda D, Santoro J (2014) New 13C-detected experiments for the assignment of intrinsically disordered proteins. J Biomol NMR 59:43–50

Pérez Y, Gairí M, Pons M, Bernadò P (2009) Structural characterization of the natively unfolded N-terminal domain of human c-Src kinase: insights into the role of phosphorylation of the unique domain. J Mol Biol 391:136–148

Pervushin K, Riek R, Wider G, Wüthrich K (1997) Attenuated T-2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. Proc Natl Acad Sci USA 94:12366–12371

Pervushin K, Vogeli B, Eletsky A (2002) Longitudinal (1)H relaxation optimization in TROSY NMR spectroscopy. J Am Chem Soc 124:12898–12902

Sattler M, Schleucher J, Griesinger C (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. Progr NMR Spectrosc 34:93–158

Schanda P (2009) Fast-pulsing longitudinal relaxation optimized techniques: enriching the toolbox. Prog NMR Spectrosc 55:238–265

Schanda P, Brutscher B (2005) Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. J Am Chem Soc 127:8014–8015

Schanda P, Kupce E, Brutscher B (2005) SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. J Biomol NMR 33:199–211

Schanda P, Van Melckebeke H, Brutscher B (2006) Speeding up three-dimensional protein NMR experiments to a few minutes. J Am Chem Soc 128:9042–9043

Shaka AJ, Barker PB, Freeman R (1985) Computer-optimized decoupling scheme for wideband applications and low-level operation. J Magn Reson 64:547–552

Shaka AJ, Lee CJ, Pines A (1988) Iterative schemes for bilinear operators; application to spin decoupling. J Magn Reson 77:274–293

Skora L, Becker S, Zweckstetter M (2010) Molten globule precursor states are conformationally correlated to amyloid fibrils of human beta-2-microglobulin. J Am Chem Soc 132:9223–9225

Smith MA, Hu H, Shaka AJ (2001) Improved broadband inversion performance for NMR in liquids. J Magn Reson 151:269–283

Solyom Z, Schwarten M, Geist L, Konrat R, Willbold D, Brutscher B (2013) BEST-TROSY experiments for time-efficient sequential resonance assignment of large disordered proteins. J Biomol NMR 55:311–321

Stanek J, Augustyniak R, Koźmiński W (2012) Suppression of sampling artefacts in high-resolution four-dimensional NMR spectra using signal separation algorithm. J Magn Reson 214:91–102

Tompa P (2009) Structure and function of intrinsically disordered proteins. CRC Press, Boca Raton

Tompa P (2012) Intrinsically disordered proteins: a 10-year recap. Trends Biochem Sci 37:509–516

Uversky VN (2013a) A decade and a half of protein intrinsic disorder: biology still waits for physics. Protein Sci 22:693–724

Uversky VN (2013b) Multitude of binding modes attainable by intrinsically disordered proteins: a portrait gallery of disorder-based complexes. Chem Soc Rev 40:1623–1634

Uversky V, Dunker AK (2013) The case for intrinsically disordered proteins playing contributory roles in molecular recognition without a stable 3D structure. F1000 Biol Rep 5:1

Uversky VN, Gillespie JR, Fink AL (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins Struct Funct Genet 41:415–427

Wright PE, Dyson HJ (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. J Mol Biol 293:321–331

Zawadzka-Kazimierczuk A, Koźmiński W, Billeter M (2012a) TSAR: a program for automatic resonance assignment using 2D cross-sections of high dimensionality, high-resolution spectra. J Biomol NMR 54:81–95

Zawadzka-Kazimierczuk A, Koźmiński W, Sanderová H, Krásný L (2012b) High dimensional and high resolution pulse sequences for backbone resonance assignment of intrinsically disordered proteins. J Biomol NMR 52:329–337

“CON-CON” assignment strategy for highly flexible intrinsically disordered proteins

Alessandro Piai[1], Tomas Hošek[1], Leonardo Gonnelli[1], Anna Zawadzka-Kazimierczuk[2], Wiktor Koźmiński[2], Bernhard Brutscher[3], Wolfgang Bermel[4], Roberta Pierattelli[1]✉, Isabella C. Felli[1]✉

[1] CERM and Department of Chemistry, University of Florence, Via Luigi Sacconi 6, 50019 Sesto Fiorentino, Florence, Italy

[2] Faculty of Chemistry, Biological and Chemical Research Centre, University of Warsaw, Żwirki i Wigury 101, 02-089 Warsaw, Poland

[3] Institut de Biologie Structurale, Université Grenoble 1, CNRS, CEA, Rue Jules Horowitz 41, 38027 Grenoble Cedex 1, France

[4] Bruker BioSpin GmbH, Silberstreifen, 76287 Rheinstetten, Germany

SUPPLEMENTARY MATERIAL

NMR experiments

In this section the pulse schemes for the following NMR experiments are described: 5D BT-(H)NCO(CAN)CONNH, 5D BT-HN(COCAN)CONNH, 3D (HACACO)NCO(CA)HA and 5D (HACA)CON(CACO)NCO(CA)HA.

E-BURP2 and REBURP shapes of durations of 1200 and 1180 ms, respectively, were employed for ^{1}H band-selective π/2 and π flip angle pulses; BIP-750-50-20 pulse shapes of duration of 200 ms were used for broadband ^{1}H inversion. For ^{13}C band-selective π /2 and π flip angle pulses G4 (or time reversed G4) and Q3 shapes of durations of 260 and 161 ms, respectively, were used, except for the π pulses that should be band-selective on the C^{α} region (Q3, 667 ms) and for the adiabatic π pulse to invert both C' and C^{α} (smoothed Chirp 500 ms, 20% smoothing, 80 kHz sweep width, 11.3 kHz RF field strength). The ^{13}C band selective pulses on C^{α} and C' were applied at the center of each region, respectively. For the $^{1}H^{\alpha}$ experiments, decoupling of ^{1}H and ^{13}C was achieved with DIPSI-2 (3.1 kHz) and GARP-4 (4.5 kHz) sequences, respectively. All gradients employed had a smoothed square shape.

In the figures, the arrows indicate the switching of the ^{13}C carrier frequency. Narrow and wide symbols stand for 90° and 180° pulses, respectively. The pulses were applied along the *x* axis unless noted differently. The rectangles represent non-selective pulses, the round shapes represent band-selective pulses. The ^{13}C π pulses that should be highly band-selective on the C^{α} region are marked by an asterisk, while the grey round shapes represent ^{13}C adiabatic pulses. Open squares on ^{1}H channel are used to indicate BIP-720-50-20 broadband inversion pulses applied in the BEST TROSY (BT) experiments. The line denoted with PFG stands for pulsed field gradients applied along the *z*-axis. The phase cycle and the duration of the delays for these experiments are reported case-by-case afterwards in the captions of the figures describing the pulse sequences.

Instead, for the 3D BT-HNCO experiment only the experimental parameters are given.

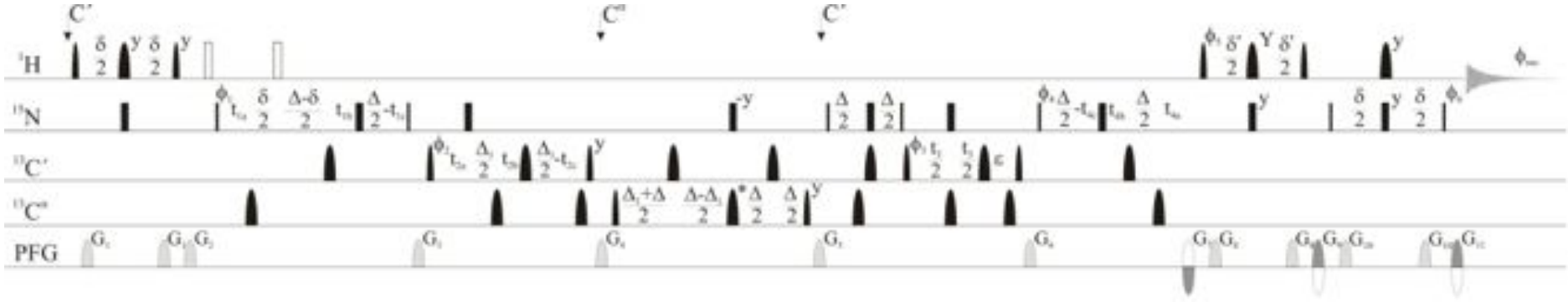


Figure S1. Pulse sequence of the 5D BT-(H)NCO(CAN)CONNH experiment

The following phase cycling was employed: ϕ_1 = 4(x), 4(-x); ϕ_2 = 2(x), 2(-x); ϕ_3 = 8(x), 8(-x); ϕ_4 = x, -x; ϕ_5 = -y; ϕ_6 = -y; ϕ_{rec} = x, 2(-x), x, -x, 2(x), 2(-x), 2(x), -x, x, 2(-x), x. The lengths of the delays were: δ = 5.5 ms, δ' = 5.5 ms $-2\cdot p_{43}\cdot 0.69$ (where p_{43} is the duration of the selective $\pi/2$ pulse on $^1H^N$), Δ = 24.8 ms, Δ_1 = 9.0 ms, $\varepsilon = t_3(0)-p_{14}$ (where p_{14} is the duration of the π pulse on $^{13}C^{\alpha}$ in the middle of the $^{13}C'$ evolution period). All gradients had 1 ms of duration, except G_1 and G_8 which were 4 ms long. Their relative strengths were the following: G_1 = 3%, G_2 = 30%, G_3 = 21%, G_4 = 17%, G_5 = 13%, G_6 = 11%, G_7 = -80%, G_8 = 5%, G_9 = 30%, G_{10} = 45%, G_{11} = 30.13% (with 100% corresponding to 0.65 T/m). The semi-constant time mode for the evolution of chemical shift was used in the first carbonyl and in both nitrogen evolution periods, whereas real time evolution was adopted for the second carbonyl evolution period. For the semi-constant time mode, the delays were set as follows: $t_{1a} = t_1/2$; $t_{1b} = t_{1a}-t_{1c}$; $t_{1c} = (\Delta/N_1)$, with N_1 the number of total increments in the t_1 dimension; $t_{2a} = t_2/2$; $t_{2b} = t_{2a}-t_{2c}$; $t_{2c} = (\Delta_1/N_2)$, with N_2 the number of total increments in the t_2 dimension; $t_{4a} = t_4/2$; $t_{4b} = t_{4a}-t_{4c}$; $t_{4c} = (\Delta/N_4)$, with N_4 the number of total increments in the t_4 dimension. Quadrature detection in the indirect dimensions was achieved by incrementing phases ϕ_1, ϕ_2 and ϕ_3 in a States manner; for quadrature detection in t_4, echo-antiecho data were recorded by inverting the sign of phases ϕ_5 and ϕ_6 together with multiplying gradients G_7, G_9 and G_{11}, respectively, by 0.8750, 1.0000, 0.6595 and, alternatively, 1.0000, 0.6667, 1.0000.

Experimental parameters used. All selective ^1H pulses were centered at 9.0 ppm. The 5D BT-(H)NCO(CAN)CONNH experiment was recorded in 11 hours with 4 scans per increment, an inter-scan delay of 0.2 s, spectral widths of 13300 (aq) × 2600 (^{15}N) × 1800 ($^{13}C'$) Hz with 1024 complex points in the acquisition dimension and 1200 hypercomplex points randomly distributed in the indirectly detected dimensions (data points density of 0.0019 %). The maximum evolution times in the indirectly detected dimensions were set to 40.4 ms for the ^{15}N dimensions (105 complex points) and to 41.7 ms for the ^{13}C dimensions (75 complex points).

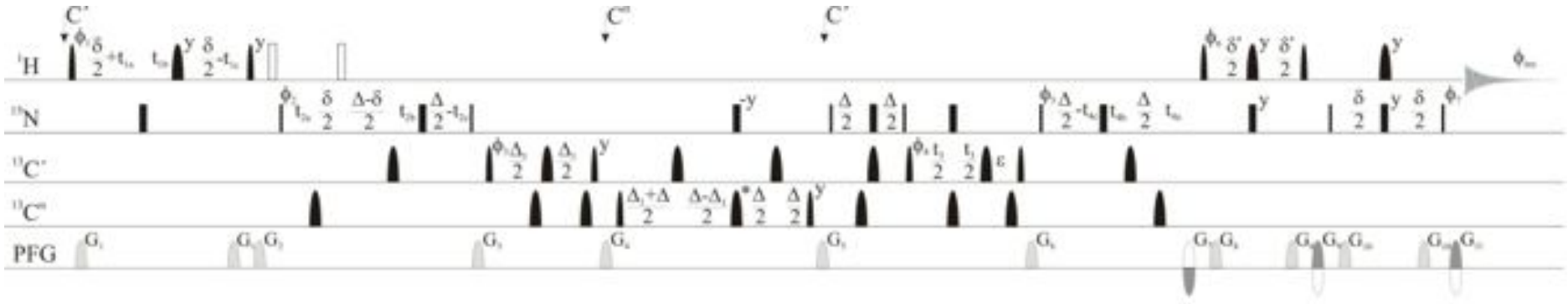


Figure S2. Pulse sequence of the 5D BT-HN(COCAN)CONNH experiment

The following phase cycling was employed: ϕ_1 = x; ϕ_2 = 4(x), 4(-x); ϕ_3 = 2(x), 2(-x); ϕ_4 = 8(x), 8(-x); ϕ_5 = x, -x; ϕ_6 = -y; ϕ_7 = -y; ϕ_{rec} = x, 2(-x), x, -x, 2(x), 2(-x), 2(x), -x, x, 2(-x), x. The lengths of the delays were: δ = 5.5 ms, δ' = 5.5 ms $-2\cdot p_{43}\cdot 0.69$ (where p_{43} is the duration of the selective $\pi/2$ pulse on $^1H^N$), Δ = 24.8 ms, Δ_1 = 9.0 ms, $\varepsilon = t_3(0)-p_{14}$ (where p_{14} is the duration of the π pulse on $^{13}C^{\alpha}$ in the middle of the ^{13}C' evolution period). All gradients had 1 ms of duration, except G_1 and G_8 which were 4 ms long. Their relative strengths were the following: G_1 = 3%, G_2 = 30%, G_3 = 21%, G_4 = 17%, G_5 = 13%, G_6 = 11%, G_7 = -80%, G_8 = 5%, G_9 = 30%, G_{10} = 45%, G_{11} = 30.13% (with 100% corresponding to 0.65 T/m). The semi-constant time mode for the evolution of chemical shift was used in the first carbonyl and in both nitrogen evolution periods, whereas real time evolution was adopted for the second carbonyl evolution period. For the semi-constant time mode, the delays were set as follows: $t_{1a} = t_1/2$; $t_{1b} = t_{1a}-t_{1c}$; $t_{1c} = (\delta/N_1)$, with N_1 the number of total increments in the t_1 dimension; $t_{2a} = t_2/2$; $t_{2b} = t_{2a}-t_{2c}$; $t_{2c} = (\Delta/N_2)$, with N_2 the number of total increments in the t_2 dimension; $t_{4a} = t_4/2$; $t_{4b} = t_{4a}-t_{4c}$; $t_{4c} = (\Delta/N_4)$, with N_4 the number of total increments in the t_4 dimension. Quadrature detection in the indirect dimensions was achieved by incrementing phases ϕ_1, ϕ_2 and ϕ_4 in a States manner; for quadrature detection in t_4, echo-antiecho data were recorded by inverting the sign of phases ϕ_6 and ϕ_7 together with multiplying gradients G_7, G_9 and G_{11}, respectively, by 0.8750, 1.0000, 0.6595 and, alternatively, 1.0000, 0.6667, 1.0000.

Experimental parameters used. All selective ^{1}H pulses were centered at 9.0 ppm. The 5D BT-HN(COCAN)CONNH experiment was recorded in 11 hours with 4 scans per increment, an inter-scan delay of 0.2 s, spectral widths of 13300 (aq) × 2600 (^{15}N) × 1800 (^{13}C') × 2000 (^{1}H) Hz with 1024 complex points in the acquisition dimension and 1200 hypercomplex points randomly distributed in the indirectly detected dimensions (data points density of 0.0018 %). The maximum evolution times in the indirectly detected dimensions were set to 40.0 ms for the ^{1}H dimension (80 complex points), to 40.4 ms for both the ^{15}N dimensions (105 complex points) and to 41.7 ms for the ^{13}C dimension (75 complex points).

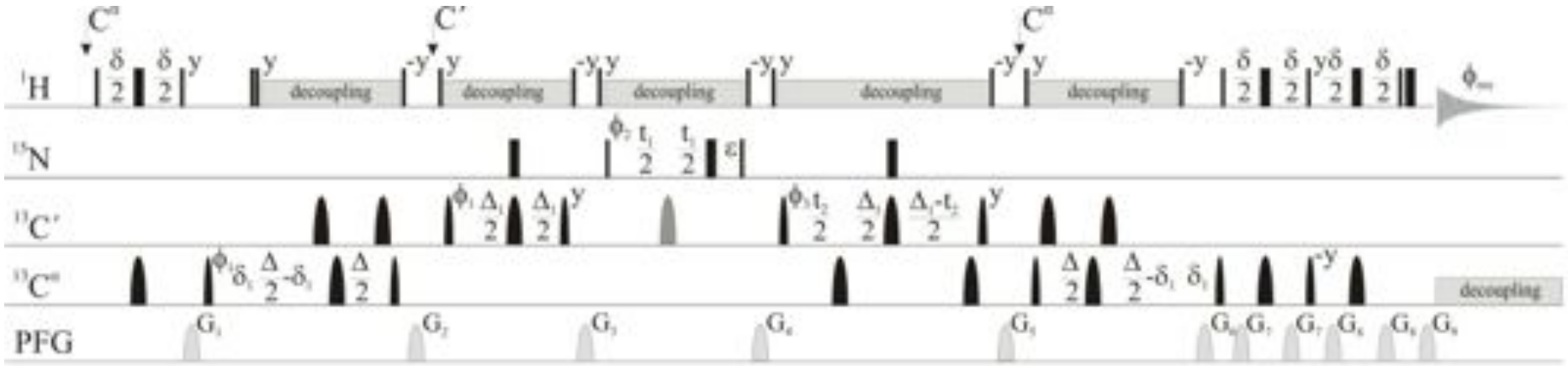


Figure S3. Pulse sequence of the 3D (HACACO)NCO(CA)HA experiment

The following phase cycling was employed: $\phi_1 = 4(x), 4(-x)$; $\phi_2 = 2(x), 2(-x)$; $\phi_3 = x, -x$; $\phi_{rec} = x, 2(-x), x, -x, 2(x), -x$. The lengths of the delays were: $\delta = 3.4$ ms, $\delta_1 = 2.2$ ms, $\Delta = 9.0$ ms, $\Delta_1 = 33.2$ ms, $\varepsilon = t_3(0)+p_8$, where p_8 is the duration of the adiabatic pulse on ^{13}C in the middle of the ^{15}N evolution period. All gradients had 1 ms of duration, except G_7 and G_8 which were 0.5 ms long. Their relative strengths were the following: $G_1 = 40\%$, $G_2 = 40\%$, $G_3 = 40\%$, $G_4 = 40\%$, $G_5 = 40\%$, $G_6 = 80\%$, $G_7 = 5\%$, $G_8 = 2\%$, $G_9 = 20.1\%$ (with 100% corresponding to 0.65 T/m). The real time mode for the evolution of chemical shift evolution was adopted for the nitrogen evolution period, whereas the constant time mode was used in the carbonyl evolution period (if necessary, the evolution of the chemical shift can be switched to the semi-constant time mode to further increase the resolution in the resulting spectrum). Quadrature detection in the indirect dimensions was achieved by incrementing phases ϕ_2 and ϕ_3 in a States manner.

Experimental parameters used. The 3D (HACACO)NCO(CA)HA experiment was recorded in 8 hours with 4 scans per increment, an inter-scan delay of 1.0 s, spectral widths of 13300 (aq) × 3500 (^{15}N) × 2000 ($^{13}C'$) Hz with 1024 complex points in the acquisition dimension and 1450 hypercomplex points randomly distributed in the indirectly detected dimensions (data points density of 20.3 %). The maximum evolution times in the indirectly detected dimensions were set to 31.4 ms for the ^{15}N dimension (110 complex points) and to 32.5 ms for the ^{13}C dimension (65 complex points)

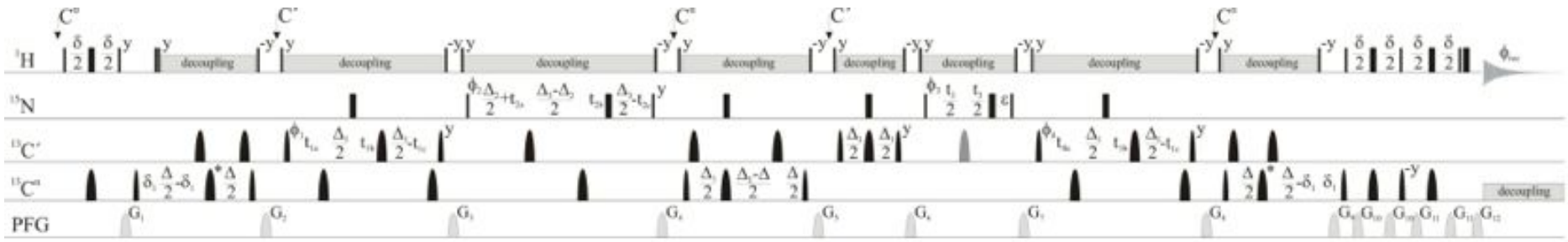


Figure S4. Pulse sequence of the 5D (HACA)CON(CACO)NCO(CA)HA experiment

The following phase cycling was employed: ϕ_1 = x, -x; ϕ_2 = 2(x), 2(-x); ϕ_3 = 4(x), 4(-x); ϕ_4 = 8(x), 8(-x); ϕ_{rec} = x, 2(-x), x, -x, 2(x), 2(-x), 2(x), -x, x, 2(-x), x. The lengths of the delays were: δ = 3.4 ms, δ_1 = 2.2 ms, Δ = 9.0 ms, Δ_1 = 28.0 ms, Δ_2 = 33.2 ms, Δ_3 = 50.0 ms, $\varepsilon = t_3(0)+p_8$, where p_8 is the duration of the adiabatic pulse on ^{13}C in the middle of the ^{15}N evolution period. All gradients had 1 ms of duration, except G_{10} and G_{11} which were 0.5 ms long. Their relative strengths were the following: G_1 = 70%, G_2 = 70%, G_3 = 70%, G_4 = 70%, G_5 = 70%, G_6 = 70%, G_7 = 70%, G_8 = 70%, G_9 = 80%, G_{10} = 5%, G_{11} = 2%, G_{12} = 20.1% (with 100% corresponding to 0.65 T/m). The semi-constant time mode for the evolution of chemical shift was used in the first nitrogen and in both carbonyl evolution periods, whereas real time evolution was adopted for the second nitrogen evolution period. For the semi-constant time mode, the delays were set as follows: $t_{1a} = t_1/2$; $t_{1b} = t_{1a}-t_{1c}$; $t_{1c} = (\Delta_1/N_1)$, with N_1 the number of total increments in the t_1 dimension; $t_{2a} = t_2/2$; $t_{2b} = t_{2a}-t_{2c}$; $t_{2c} = (\Delta_3/N_2)$, with N_2 the number of total increments in the t_2 dimension; $t_{4a} = t_4/2$; $t_{4b} = t_{4a}-t_{4c}$; $t_{4c} = (\Delta_1/N_4)$, with N_4 the number of total increments in the t_4 dimension. Quadrature detection in the indirect dimensions was achieved by incrementing phases ϕ_1, ϕ_2, ϕ_3 and ϕ_4 in a States manner.

Experimental parameters used. The 5D (HACA)CON(CACO)NCO(CA)HA experiment was recorded in 2 days and 7 hours with 8 scans per increment, an inter-scan delay of 1.0 s, spectral widths of 13300 (aq) × 3500 (^{15}N) × 2000 (^{13}C') Hz with 1024 complex points in the acquisition dimension and 1150 hypercomplex points randomly distributed in the indirectly detected dimensions (data points density of 0.015 %). The maximum evolution times in the indirectly detected dimensions were set to 35.7 ms for the ^{15}N dimensions (125 complex points) and to 35.0 ms for the ^{13}C dimensions (70 complex points).

3D BT-HNCO experiment

The Bruker b_trhncogp3d pulse program was used. The nitrogen chemical shift evolution was extended from the constant time to the semi-constant time mode.

Experimental parameters used. The 3D BT-HNCO experiment was recorded in 3 hours with 4 scans per increment, an inter-scan delay of 0.2 s, spectral widths of 13300 (aq) × 2600 (^{15}N) × 1800 (^{13}C') Hz with 1024 complex points in the acquisition dimension and 1575 hypercomplex points randomly distributed in the indirectly detected dimensions (data points density of 20.0 %). The maximum evolution times in the indirectly detected dimensions were set to 40.4 ms for the ^{15}N dimension (105 complex points) and to 41.7 ms for the ^{13}C dimension (75 complex points).

5.1.4

NMR methods for the study of IDP structure, dynamics, and interactions: general overview and practical guidelines

(Book chapter)

Bernhard Brutscher[a], Isabella C. Felli[b], Sergio Gil-Caballero[c], Tomáš Hošek[b], Rainer Kümmerle[c], Alessandro Piai[b], Roberta Pierattelli[b], Zsófia Sólyom[a]

[a] *Institut de Biologie Structurale, Université Grenoble 1, CNRS, CEA, 71 avenue des Martyrs, 38044 Grenoble Cedex 9, France*

[b] *CERM and Department of Chemistry Ugo Schiff, University of Florence, via Luigi Sacconi 6, Sesto Fiorentino, 50019, Italy*

[c] *Bruker BioSpin AG, Industriestrasse 26, Fällanden, 8117, Switzerland*

Contents

1 A short introduction to NMR spectroscopy and instrumentation 110

1.1 The basic principles of NMR 110

1.2 Recent improvements in NMR instrumentation 113

2 The effect of order and disorder on NMR observables 114

2.1 Structured versus disordered proteins 114

2.2 NMR peak positions and chemical shifts 115

2.3 Secondary chemical shifts 116

2.4 Line splittings and spin coupling constants 117

2.5 NMR spin relaxation, line widths and intensities 118

2.6 ^{15}N relaxation parameters 120

3 Particular challenges and bottlenecks for NMR studies of IDPs 122

3.1 Spectral resolution 122

3.2 Experimental sensitivity 123

3.3 Experimental time requirements 123

3.4 Sample optimization 124

3.5 Proline residues are abundant in IDPs 124

3.6 IDP studies at near-physiological conditions (pH and T) 125

4 2D HN and CON NMR experiments: the fingerprint of an IDP 126

4.1 NMR properties of ^{1}H, ^{13}C and ^{15}N in IDPs 126

4.2 2D ^{1}H-^{15}N and ^{13}C'-^{15}N correlation fingerprint spectra of IDPs 127

5 NMR tools for overcoming major bottlenecks of IDP studies 130

5.1 Multidimensional NMR, indirect frequency editing and non-uniform sampling 130

5.1.1 Real-time, constant-time and semi-constant time frequency editing . 130

5.1.2 Strategies for non-uniform data sampling in multi-D NMR 131

5.2 Selecting ^{15}N spin states with favorable transverse relaxation properties . . 133

5.3 Longitudinal-relaxation enhancement for increased sensitivity and reduced acquisition times 134

5.4 ^{13}C detected experiments 138

5.4.1 Homonuclear ^{13}C decoupling 138

5.4.2 Starting polarization source 140

6 From 2D to 3D: from simple snapshots to site-resolved characterization of IDPs **142**

6.1 Sequential NMR assignment: the general strategy 142

6.2 Sequential NMR assignment: 3D experiments 143

6.2.1 $^{1}H^{N}$ detected experiments 143

6.2.2 ^{13}C detected experiments 145

6.2.3 Aliphatic ^{1}H detected experiments 147

6.3 Experiments for amino acid type editing or selection 147

6.4 Automated assignment 149

7 High-dimensional NMR experiments (nD, with n > 3) **151**

7.1 Analysis and inspection of high-dimensional NMR spectra 151

7.2 Examples of high-dimensional experiments tailored for IDPs 152

8 Conclusions and perspectives **156**

Appendix **158**

Tables **163**

Figures **164**

References **191**

1 A short introduction to NMR spectroscopy and instrumentation

Nowadays, the possibilities offered by modern NMR spectroscopy are enormous, and cover a wide range of applications in physics, chemistry, biology, medicine and material sciences. Past and present progress in the field is mainly based on technical improvements of the NMR spectrometer (magnetic field strength, NMR electronics, probes,...) and on the development of a variety of pulse sequences exploiting in an ingenious way the basic principles of NMR spectroscopy, in order to obtain the desired information on a particular sample. Therefore, the first part of this chapter will be dedicated to the most recent key developments in NMR instrumentation, mainly stimulated by emerging scientific challenges, and to the basic principles of NMR spectroscopy.

The phenomenon of nuclear magnetic resonance (NMR) was discovered by Purcell et al. [1] and Bloch [2], and shortly afterwards they were awarded the Nobel Prize in Physics *"for their development of new methods for nuclear magnetic precision measurements and discoveries in connection therewith"*. Since its discovery, NMR spectroscopy, further to its fundamental role in physics, has become a very powerful tool in chemistry and biology for structural studies of small organic and inorganic compounds as well as large systems, including polymers and biomolecules. Besides X-ray crystallography and electron microscopy, NMR is the only method capable of providing atomic resolution information on the structure of biological macromolecules. Compared to X-ray crystallography, it allows studies of biological macromolecules in the liquid state - meaning that crystallization of the molecule is not required. This makes possible structural studies of highly dynamic systems such as intrinsically disordered proteins (IDPs). NMR spectroscopy can also be employed for studies of interactions of biomolecules with small ligands, metal ions as well as with other biological macromolecules. Recently, the importance of regarding biomolecules as dynamic ensembles instead of a single static entity has become widely recognized. NMR spectroscopy provides a unique tool to access dynamics information at atomic resolution, from the picosecond timescale to slow exchange processes on the second (or even slower) timescale. This is especially important for intrinsically disordered proteins that are often characterized by heterogeneous dynamic properties along the polypeptide chain.

1.1 The basic principles of NMR

NMR spectroscopy studies the interaction of matter with radiofrequency electromagnetic waves that excite magnetic transitions of the atomic nuclear spins. Indeed an atomic

nucleus, in order to be observable by NMR spectroscopy, has to possess a nonzero spin quantum number I. For NMR studies of proteins the relevant nuclear isotopes are ^{1}H, ^{13}C, and ^{15}N, all of them have a spin quantum number of $I = 1/2$. When the nuclei are placed in a static magnetic field B_0 that is by convention aligned along the z-axis, the nuclear magnetic momentum μ of the spins interact with the magnetic field leading to a splitting of the energy levels (Zeeman splitting) according to equation:

$$E = -\boldsymbol{\mu} \cdot \boldsymbol{B} = -\mu_z B_0 = -\gamma I_z B_0 = -\gamma m \hbar B_0 \tag{1}$$

where γ is the gyromagnetic ratio and $\hbar$ is the reduced Planck constant. For spin 1/2 nuclei the magnetic quantum number m can take the values of $m = +1/2$ and $m = -1/2$. Gyromagnetic ratios and natural abundance of the nuclear isotopes important in biomolecular NMR are summarized in Table 1.1.

Equation (1) is of utmost importance for the sensitivity of NMR experiments. The energy difference between the two Zeeman energy levels is given by:

$$\Delta E = -\gamma \hbar B_0 = \hbar \omega_0 \tag{2}$$

with ω_0 the characteristic NMR frequency (Larmor frequency) of a nuclear spin at a given magnetic field strength.

At thermal equilibrium and temperature T, the ratio of nuclear spins in the lower (E_α) and in the higher energy state (E_β) can be calculated from the Boltzmann distribution:

$$\frac{N_\beta}{N_\alpha} = e^{-\frac{\Delta E}{kT}} \tag{3}$$

where k is the Boltzmann constant.

The NMR signal is proportional to the magnetization that in turn is dependent on the spin polarization P that is given by the population difference between the two states divided by the total number of spins:

$$P = \frac{N_\alpha - N_\beta}{N_\alpha + N_\beta} \tag{4}$$

In case of ^{1}H nuclei, that possess the highest gyromagnetic ratio among the spin $\frac{1}{2}$ nuclei in proteins, the ratio of spins in the upper energy state versus the lower energy state is 0.999872 at room temperature on a 800 MHz magnet (B_0 = 18.8 T). This means that only a very small fraction of the spins present in the sample contributes to the observed NMR signal. Furthermore, in the case of ^{13}C and ^{15}N nuclei, often called heteronuclei, the

natural abundance of the NMR active nuclei is very low, as shown in Table 1.1. Therefore isotope enrichment techniques have been developed to enhance the sensitivity of NMR techniques involving such heteronuclei.

Sensitivity and resolution are the two most important factors influencing the outcome of NMR experiments. NMR sensitivity, defined as the signal to noise ratio (SNR) obtained in a fixed amount of time, is given by the following equation:

$$SNR \propto \frac{N}{V}\gamma_{exc}\gamma_{det}^{3/2}B_0^{3/2}n_{scan}^{1/2}\frac{1}{\sqrt{R_s(T_a+T_s)+R_c(T_a+T_c)}} \tag{5}$$

where N is the number of spins, V the active volume, γ_{exc} the gyromagnetic ratio of the excited nuclei and γ_{det} that of the detected nuclei, B_0 the static magnetic field, n_{scan} the number of scans, R_S the resistance of the sample and R_c that of the coil, T_a the temperature of the preamplifier and T_s and T_c those of the sample and of the coil, respectively. [3,4] Other factors that depend on the properties of the sample under investigation and on the type of pulse sequence used contribute to the sensitivity of an NMR experiment, and thus to the possibility to access the desired information. These will be discussed in detail in the following paragraphs.

Some general conclusions for improving the sensitivity of an NMR experiment can be derived from equation (5). The NMR signal is proportional to the amount of spins present in the sample. Thus it is desirable to use highly concentrated samples for NMR spectroscopy. However, in practice, the maximum concentration of a protein sample is often limited by protein solubility and obtaining large quantities of isotopically labeled proteins can be expensive and time consuming. Most protein NMR experiments use proton excitation and detection because of the high gyromagnetic ratio of these nuclei, nevertheless ^{13}C direct detection also provides a valuable tool for biomolecular NMR applications (see paragraph 5.4). The linear dependence on B_0 explains the ongoing efforts of NMR manufacturers to develop magnets with higher magnetic fields. To date, the highest magnetic field of a commercial magnet corresponds to 1 GHz proton Larmor frequency (23.5 T), and 1.2 GHz magnets (28.2 T) are under development. The SNR increases with the square root of the number of scans (experimental repetitions). Increasing the overall measurement time thus presents a common way for spectral improvement. The last contribution to NMR sensitivity comes from the electronic detection circuit. The SNR of NMR increases with lowering the temperature of the preamplifier (T_a), the sample (T_s), and the coil (T_c), as well as the resistance of the sample (R_s) and the coil (R_c). A recent development that greatly enhanced the sensitivity of NMR spectroscopy was the introduction of probes with

cryogenically cooled (to about 20 K) detection circuits. However, protein samples with high ionic strengths degrade the beneficial effects of cryogenically cooled probes.

Resolution is the other important factor to extract atomic resolution information from NMR spectra. It depends on chemical shift dispersion and signals linewidths. Two peaks can be resolved when their difference in frequency is larger with respect to their linewidths. Nuclear spins that are characterized by favorable chemical shift dispersion and small linewidths for the system investigated should thus be exploited to improve the resolution of the spectra. The possibility to introduce additional indirect dimensions in NMR experiments provides the other invaluable tool to enhance the resolution by spreading signals in additional dimensions and thus reducing the possibility of accidental cross-peak overlap.

1.2 Recent improvements in NMR instrumentation

To be soon provided by Bruker.

2 The effect of order and disorder on NMR observables

In the following we will introduce the main NMR observables that are used to derive structural and dynamic information on proteins, with a particular emphasis on how the particular features of IDPs affect these NMR observables.

2.1 Structured versus disordered proteins

In order to better understand the impact of structural order and disorder on the NMR experiments and spectral parameters, it is important to recognize the principal features that distinguish a well-folded globular protein from a largely unfolded, highly flexible IDP. These are illustrated in Figure 2.1.

Hydrophobic interactions and the formation of hydrogen bonds are the main driving forces for the folding process of a polypeptide chain into a globular 3D structure, and for the stabilization of such a compact fold. The main consequences of the presence of a stable, well-defined structure, that are of importance for the NMR properties of the molecule are a high proton density, the exclusion (to a large extent) of water from the interior of the protein and, to a first rough approximation, the possibility to describe its rotational motion with an overall rotational correlation time. The high proton density and the large rotational correlation time makes the ^{1}H-^{1}H NOEs the main source of structural information for globular proteins, whereas they are only of little use in IDP research, where the effect is mainly restricted to proton pairs that are close in the primary sequence and is on average smaller than for globular proteins. Indeed, the dynamic behavior of IDPs is significantly different from that of structured proteins, as a result of the small energy difference between conformers that is responsible for the high flexibility typical of IDPs and allows easy inter-conversion between many different conformations. Therefore, even to a first approximation, a single overall rotational correlation time cannot be defined for IDPs and local effective correlation times are on average smaller with respect to those used to describe structured proteins of similar size. The differences in ^{1}H density and molecular dynamics between IDPs and globular proteins also have a strong influence on the NMR signals chemical shifts and the relaxation parameters. Conformational averaging drastically reduces contributions to chemical shifts deriving from the local environment and causes severe resonance overlap. The high flexibility of IDPs has a strong impact on nuclear relaxation rates, and thus on NMR linewidths. Depending on the time scale of motions, several situations can be encountered: on one hand, conformational exchange processes can result in extensive line broadening, and in the extreme case in the absence

of any detected NMR signal. On the other hand fast motions result in narrow NMR lines, a property that makes highly flexible IDPs particularly amenable to NMR characterization. For such highly flexible IDPs a variety of complex NMR experiments can be conceived, where the favorable relaxation properties allow for multiple transfer steps and the narrow linewidths ensure high resolution in the resulting spectra. Finally, IDPs are also characterized by non-compact conformations and the exposure of labile protons, e.g. amide protons, to the solvent results in high exchange rates. In fact, the measurement of solvent exchange rates is used to distinguish between highly structured and disordered regions of proteins by identifying solvent exposed and solvent protected amide sites [5]. Approaching physiological pH and temperature, exchange rates of solvent exposed amide protons may broaden resonances beyond detection and in this case alternative nuclear spins, such as aliphatic ^{1}H or ^{13}C, should be detected to access information on the IDP through NMR [6].

2.2 NMR peak positions and chemical shifts

Peak positions are dependent on the resonance frequencies of the observed nuclei. Based on equation (2) one would expect to observe a single line in the NMR spectrum for each nucleus of a given type (^{1}H, ^{13}C, ^{15}N) at the respective Larmor frequency. However, this would be true only in the hypothetical experiment of observing "naked" nuclei. In practice, nuclei of the same type resonate at slightly distinct frequencies if they are in chemically different environments within the molecule. The reason is that the nuclei experience a net magnetic field B that is the sum of the static field B_0 and secondary shielding fields induced by the local electronic environment:

$$B = (1 - \sigma)B_0 \tag{6}$$

where σ is the isotropic average shielding factor. This is at the origin of the term "chemical shift" used to measure resonance positions in NMR spectroscopy. Chemical shift values δ are typically measured relative to the chemical shift of a standard according to:

$$\delta = \frac{\Omega - \Omega_{ref}}{\omega_0} \times 10^6 \tag{7}$$

with ω_0 the Larmor frequency in MHz, Ω the frequency offset of the nucleus of interest in Hz and Ω_{ref} the offset of the standard in Hz. In this way, the chemical shift becomes a field-independent quantity, expressed in ppm (*parts per million*). The chemical shift

makes NMR an atomic resolution technique: once the resonances are assigned to their respective nuclei, their response to further manipulations can be followed.

Chemical shifts can be very characteristic for nuclei in different chemical moieties, such as the different amino acids constituting a protein, and they provide the first useful information that can be obtained from the NMR spectrum. The stable local three-dimensional structure, or partially populated secondary structural conformations, create a unique local electronic environment, and thus the contribution to chemical shift coming from isotropic shielding is different for each nuclear spin. In IDPs, the lack of a stable structure results in averaging of a large part of the contributions to the chemical shift coming from the local chemical environment, which is the reason for the low chemical shift dispersion observed in the NMR spectra (Figure 2.2).

In addition, peak intensities (areas/volumes) can be easily determined and may provide useful atomic resolution information. Peak intensities in the NMR spectra are proportional to the number of nuclear spins giving rise to the peaks. They also report on differential signal loss during an NMR experiment due to relaxation processes. Therefore they provide a first useful indication of heterogeneous structural and dynamic properties of a protein. Peak linewidths would also provide useful information but they are more difficult to measure as they are influenced by different factors (see paragraph 2.5). In many cases, for simplicity, the determination of peak heights is used.

2.3 Secondary chemical shifts

The chemical shifts of backbone nuclei are sensitive to the local backbone geometry (dihedral angles), and therefore provide useful information on the occurrence and propensity of secondary structural elements along the protein backbone [7,8]. For a ^{13}C, ^{15}N enriched protein, a large number of backbone and side-chain chemical shifts ($^{1}H^{N}$, ^{15}N, $^{13}C'$, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, ...) can be measured, all of which are indicative of secondary structure even if to a different extent. In order to extract this information, a so-called secondary chemical shift is computed as the difference of the measured chemical shift and a predicted random coil value for each nucleus [9–14]. Random coil chemical shifts are the theoretical chemical shifts of a random coil polypeptide of the same amino acid sequence, characterized by lack of long-range order and secondary structure. However, the conformational sampling of a polypeptide is never completely random, in the sense that all dihedral angles are sampled with equal probability, because of steric hindering and chemical interactions between neighboring side-chains. The distribution of the sampled dihedral angles along the protein backbone is determined by the conformational Gibbs free energy that in turn depends

on pH, temperature and solvent effects. The accuracy of secondary chemical shifts thus depends on the quality of the predicted random coil chemical shift values. This aspect becomes even more crucial for IDPs, since the measured chemical shifts are usually close to the corresponding random coil values. Figure 2.3 illustrates how C^{α} and C^{β} secondary chemical shifts can be used to identify α-helical and β-sheet regions and highly unstructured segments in globular proteins, as well as residual secondary structures in IDPs.

Different tables of ^{13}C and ^{15}N random coil values for different amino acids, taking into account next and previous neighbors, pH and solvent effects can be found in the literature [11–15]. One of the most commonly used random coil chemical shift data sets, based on chemical shift measurements on small peptides, is the one reported by Wishart et al. [16]. Recently, temperature and pH correction factors have been introduced from a peptide-based study where glutamine peptides were used instead of taking poly-glycine peptides (Wishart et al.), as their conformational sampling is considered to be more representative [12]. The other type of reference data can be obtained from protein chemical shifts databases. This approach has the inconvenience that sample conditions such as pH and temperature vary between different entries. Furthermore, neighboring and other chemical effects will be biased depending on the composition of the entries. The most commonly used database is refDB by Zhang et al. [15]. Recently, an IDP-based random coil chemical shift database has been introduced [12,14,17]. Neighbor corrections are usually required to account for changes in conformational sampling because of steric clashes and other chemical effects such as electrostatics or ring current shifts [11,16,17].

2.4 Line splittings and spin coupling constants

Nuclear spins are never completely isolated. Thus mutual interactions between neighboring nuclear spins are always present and give rise to couplings. The coupling derives from two main contributions: scalar (through bonds) and dipolar (through space). While the latter is averaged out in isotropic solutions, the former, mediated by the electrons in covalent bonds (scalar or ^{n}J-coupling, where n indicates the number of covalent bonds separating the two nuclei involved in the coupling), causes splitting of the NMR signal. The frequency differences between the multiplet components of an NMR signal measured in Hz then reflect the strength of the scalar couplings. Scalar couplings involving ^{1}H, ^{13}C and ^{15}N are exploited for coherence transfer in the majority of multidimensional NMR experiments discussed in the following paragraphs. Scalar couplings, in particular ^{3}J-couplings in a polypeptide, also contain valuable structural information as they depend on the intervening dihedral angle according to the Karplus relation [18]. Inter-conversion

between different conformers typical of IDPs results in averaged scalar couplings; while this has little impact on coherence transfer in multidimensional NMR experiments, it does influence the interpretation of ^3J-couplings in terms of intervening dihedral angles. For example, $^3J_{\mathrm{HH}}$ couplings are sometimes used to confirm the presence of partially populated secondary structural elements [19–22].

The magnetic moments of two nuclear spins interact "through space" via the dipolar mechanism with an interaction strength that is inversely proportional to r^3, with r the inter-nuclear distance. The dipolar interaction has a particular angular dependence with respect to the static magnetic B_0 field and, as a consequence, in isotropic solutions where all orientations are sampled with equivalent probability this interaction averages to zero and does not give rise to line splittings. However, it is possible to reintroduce the dipolar coupling by either exploiting the natural magnetic anisotropy of a molecule, resulting in slightly unequal sampling of the molecular orientations in a strong static magnetic field (partial molecular alignment) [23–25], or by dissolving the protein in an anisotropic "alignment" medium, e.g. a liquid crystalline solution [26]. This allows residual dipolar coupling (RDC) to be measured as a line splitting, similarly to what is done for the measurement of J-coupling constants. In fact, the measured couplings in a partially aligned sample are the sum of the scalar and residual dipolar couplings. RDCs provide important information on local and global structure and dynamics in the molecule as explained in more detail in chapter by Kragelj.

2.5 NMR spin relaxation, line widths and intensities

Relaxation of nuclear spins after excitation by a radio-frequency is caused by time-dependent local magnetic fields induced by the molecular tumbling and local rotational fluctuations that modulate the anisotropic spin interactions, the chemical shielding anisotropy (CSA) and the dipolar (DD) interaction. Additional contributions to spin relaxation arise from exchange processes, e.g. chemical exchange between labile protein and solvent protons, or exchange between different molecular conformations. In a simplified representation of nuclear spin relaxation, the relaxation process is described by two time constants. The longitudinal relaxation time constant T_1 accounts for the return of the spin system to thermal equilibrium associated with a loss of energy, while the transverse relaxation time constant T_2 describes the dephasing of coherence, an adiabatic process. A main consequence of spin relaxation for a particular NMR experiment is the loss of signal during the various transfer steps and chemical shift evolution delays. In addition, the longitudinal relaxation properties of the excited spin species determine the rate at which a pulse sequence can be

repeated as will be explained in more detail in paragraph 5.3 while transverse relaxation properties of nuclear spins determine their NMR linewidth.

The NMR linewidth is given by the transverse relaxation rate R_2 (reciprocal of T_2) of the detected nuclear spin, $\Delta\nu = R_2/\pi$, plus inhomogeneous contributions arising from the experimental setup (sample heterogeneities, B_0 inhomogeneity, temperature gradients, etc.). IDPs, in particular highly flexible ones, are generally characterized by large transverse relaxation time constants T_2, which leads to narrow peaks in the NMR spectra in comparison to the ones observed for globular proteins of comparable size. On the other hand, chemical and conformational exchange processes can be highly pronounced in IDPs, causing broadening of the peaks especially in the spectra exploiting labile amide protons or in the case of structurally and dynamically heterogeneous proteins. Comparisons of cross-peak intensities (areas/volumes) thus report on differential intensity loss during the pulse sequence comparison of cross-peak heights also reports on differences in linewidth. Often, the NMR analysis software determines the peak heights in the spectrum during the procedure of peak picking. When the peak is picked the program fits a function, such as Lorentzian or Gaussian, and the maximum gives the peak height. For IDPs, peak heights in simple 2D spectra like the 2D HN and 2D CON (see paragraph 4.2), are often heterogeneous because relaxation properties of the spins vary depending on the extent of transient structure and on differences in local mobility. Therefore they provide a first indication of the different structural and dynamic properties of the protein.

Relaxation effects, including auto-correlated as well as cross-correlated relaxation rates, can be in many cases accurately quantified and used to extract structural and dynamic information. In this chapter we mainly focus on the determination of ^{15}N relaxation rates, one of the major tools to achieve information on the local dynamics of different parts of protein backbones [27, 28]. Other interesting relaxation rates that can be determined and exploited to achieve structural and dynamic information on IDPs are instead discussed in chapter by Konrat (paramagnetic relaxation enhancements, cross correlation rates). The determination of ^{1}H-^{1}H NOEs instead is not discussed in detail as it plays a minor role in the study of IDPs and many excellent books and reviews are available [29, 30]. However, the major contributions to ^{1}H relaxation are discussed in paragraph 5.3 because they are at the basis of longitudinal relaxation enhancement effects that provide a valuable tool for the design of NMR experiments.

2.6 ^{15}N relaxation parameters

The quantification of nuclear spin relaxation effects provides a valuable tool to characterize local and global molecular motions. Among the many different methods that have been proposed in the literature, ^{15}N relaxation measurements, notably ^{15}N T_1, ^{15}N T_2 and the ^{1}H-^{15}N heteronuclear NOE, are the most commonly used to characterize the dynamic behavior of proteins, both for globular proteins and IDPs. Furthermore, the major contributions to ^{15}N relaxation of a backbone amide ^{15}N in a protein are the ^{15}N CSA and the ^{15}N-^{1}H dipolar interaction with the directly bound proton, which to a good approximation, can be considered of equal magnitude throughout the protein backbone. This means that variations in ^{15}N relaxation properties of backbone amide ^{15}N spins can be interpreted in terms of differences in local motions [27]. The measured ^{15}N relaxation parameters provide quantitative information on the amount and the time scale of motion experienced by the ^{15}N-^{1}H amide group of a specific residue. ^{15}N relaxation (^{15}N T_1, ^{15}N T_2 and the ^{1}H-^{15}N heteronuclear NOE) is sensitive to motion occurring on the *pico-* to *nano*-second time scale, while the ^{15}N T_2 relaxation times are also sensitive to conformational exchange processes in the *micro-* to *milli*-second time range. As an example, Figure 2.4 shows ^{1}H-^{15}N heteronuclear NOE values measured for individual amide sites along the backbones of different proteins; being especially sensitive to fast local motions, these observables allow to easily identify protein regions characterized by different levels of local mobility.

As mentioned in the previous paragraph, spin relaxation in NMR is caused by oscillating magnetic fields induced by stochastic local and global rotational motions. In the Bloch-Wangsness-Redfield theory of nuclear spin relaxation [31–33], the frequency distribution of the molecular motions is described by the power spectral density function $J(\omega)$. The power spectral density components at zero-frequency $J(0)$ and at the different transition frequencies of the coupled ^{15}N-^{1}H spin system ($J(\omega_N)$, $J(\omega_H)$, $J(\omega_H - \omega_N)$, $J(\omega_H + \omega_N)$) contribute to ^{15}N relaxation. The dependence of the measured relaxation parameters on the power spectral density components is given by:

$$\frac{1}{T_1} = \frac{d^2}{20}[J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)] + \frac{c^2}{15}J(\omega_N)$$

$$\frac{1}{T_2} = \frac{d^2}{40}[4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 3J(\omega_H) + 6J(\omega_H + \omega_N)] + \frac{c^2}{90}[J(0) + 3J(\omega_N)] + R_{ex}$$

$$NOE = 1 + \frac{d^2}{20}\frac{\gamma_H}{\gamma_N}[6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)] \times T_1 \quad (8)$$

with $d = \frac{\mu_0}{4\pi}\gamma_H\gamma_N\hbar r_{NH}^{-3}$ the dipolar coefficient, $c = \omega_N\Delta\sigma$ the ^{15}N chemical shift anisotropy and R_{ex} the contribution from conformational or chemical exchange. For the case of an

isotropically tumbling rigid molecule with correlation time τ_c, the spectral density function is given by $J(\omega) = \frac{2\tau_c}{1+\omega^2\tau_c^2}$.

The standard analysis of ^{15}N relaxation data for globular proteins uses the model-free formalism introduced by Lipari and Szabo where the molecular tumbling is separated from local motions [34]. This separation is justified by the difference in time scales of these motions in the case of globular proteins, but this model is not rigorous for random coil like polymers, as pointed out in their paper. However, calculation of an effective local correlation time from the T_1/T_2 ratio provides a quick way of obtaining information on the local rigidity [35]. Alternatively, the reduced spectral density mapping approach can be used to extract information about the shape of the spectral density function [36]. The method exploits the near ten-fold difference in the gyromagnetic ratios of ^{1}H and ^{15}N, and substitutes $J(\omega_H)$, $J(\omega_H - \omega_N)$, $J(\omega_H + \omega_N)$ with an effective $J(0.87\omega_H)$ value. The validity of this approach for very flexible IDPs has been recently investigated and a slightly optimized protocol was proposed, including the removal of exchange contributions by measurement of additional cross-correlated relaxation rates [37].

3 Particular challenges and bottlenecks for NMR studies of IDPs

A wide range of NMR experiments has been developed throughout the years for the study of globular proteins and their interactions, with the major objective of providing high-resolution structural and dynamic information. These experiments provide the natural starting point for the NMR investigation of IDPs. However, the peculiar properties of IDPs have a strong impact on the NMR spectra, as already outlined in the previous paragraph, and thus on the NMR experiments. This means that the conventional NMR experiments need to be tailored for the specific properties of highly disordered proteins in order to study IDPs of increasing size and complexity. The main bottlenecks for NMR studies of IDPs will be briefly reviewed in the following sections.

3.1 Spectral resolution

In order to extract structural and dynamic information for single nuclear sites in an IDP, we need sufficient spectral resolution to distinguish individual resonances (or correlation peaks). As mentioned in the previous paragraph the first consequence of the lack of a stable structure is the averaging of a large part of the contributions to the chemical shift deriving from the local chemical environment. This results in a drastic reduction of the chemical shift ranges for the different nuclear spin species (^{1}H, ^{13}C, ^{15}N) and thus in a problem of strong overlap in the corresponding NMR spectra (see Figure 3.1).

This resolution problem is especially pronounced for ^{1}H NMR, as will be discussed in more detail later on. Identifying strategies to overcome the problem of resonances/cross-peaks overlap is of key importance to be able to study IDPs of increasing size and complexity. Several strategies to address this critical point are discussed in the next paragraphs (such as exploiting the favorable resolution in exclusively heteronuclear NMR, correlating nuclei of neighboring amino acids, exploiting amino acid selection to simplify spectra and identify residue types). As said, amino acid sequences of IDPs are often highly repetitive [38]. In the absence of a 3D structure, the nuclei within these repetitive elements resonate at nearly identical NMR frequencies. As a consequence, to obtain resolved NMR correlation peaks for such peptide stretches, high-dimensional ($\geq$ 4D) experiments might be required as will be explained later on.

3.2 Experimental sensitivity

We have already introduced the theoretical basis and the main determinants of NMR sensitivity in paragraph 1.1 Here we will focus on IDPs that are by definition very flexible macromolecules. From a NMR sensitivity point of view this flexibility has both advantages and disadvantages. The fast timescale molecular motions in IDPs are responsible for reduced effective rotational correlation times, a feature that in many cases contributes to long spin coherence lifetimes (long T_2) and narrow NMR lines. In principle, this enables the design of complex high-dimensional NMR pulse schemes to achieve the necessary resolution required for IDPs. It is interesting to note that for highly flexible IDPs the increase in molecular size does not have a major impact on relaxation times and linewidths, while it does for folded proteins. However, for regions with a significant amount of transient structure, this is no longer the case, as illustrated by amide ^{15}N T_2 relaxation time constants of the NS5A protein of HCV (Figure 3.2). There are peptide segments that have T_2 relaxation time constants four times shorter than what is observed for other regions. This results in a large dynamic range of peak intensities observed in the NMR spectra. This feature becomes even more pronounced for more complex NMR experiments involving an increasing number of transfer steps and frequency editing periods.

The effect on the NMR spectra of partially structured peptide regions and of extensive conformational dynamics can become so pronounced that most of the nuclear spins are affected. No or only very limited solutions remain to study by NMR proteins characterized by strongly line-broadened resonances. Examples that often fall into this category are the so-called molten-globule states that represent the"dark side" of biomolecular NMR, in between well folded globular and highly disordered states.

Sensitivity can also be reduced because of protein aggregation that is governed by the same principles as protein folding. Some IDPs, such as α-synuclein, tau, or prion proteins, are known to cause neurodegenerative diseases in the aggregated form. Tendencies for aggregation can hamper NMR studies of these proteins. On the other hand, higher molecular mass aggregates become amenable to other NMR techniques, such as solid-state MAS NMR, which however is not treated in this chapter [39–41].

3.3 Experimental time requirements

Sample stability problems are often encountered because of aggregation, as mentioned before, or proteolytic degradation of IDPs. When the stability of the protein permits sample stabilization by boiling, the proteases can be deactivated by heat denaturation.

However, in case of significant residual structure, this is often not a valuable option as extensive heat treatment may result in irreversible changes of the structural features of the IDP (chapter by Calçada). In such cases, the NMR spectra have to be acquired before sample degradation occurs. The requirements of short experimental times and at the same time high spectral resolution, achieved by long acquisition times in all dimensions of a high-dimensional (3D, 4D,...) experiment seem to be contradictory, but they can be reconciled by the use of fast NMR data acquisition techniques discussed in paragraphs 5.1 and 5.3.

3.4 Sample optimization

The conformational dynamics and transient structure of IDPs are highly sensitive to experimental conditions such as buffer composition, salt concentration, pH and temperature. Under certain conditions, some parts of the IDP may undergo conformational dynamics on a timescale that leads to extensive line broadening in the NMR spectra, and thus missing correlation peaks. Optimization of the sample conditions is thus even more important than for globular proteins. On the other hand, because of the dependence of the structural features of the IDP on the experimental conditions, one may want to study the IDP under close to physiological conditions, such as neutral pH and relatively elevated temperature. Also studies of enzymatic reactions, for example the occurrence of post-translational modifications, often require neutral pH and high temperature to ensure optimal enzyme activity. Finally, in-cell NMR can be used in order to experimentally show that IDPs remain flexible *in vivo* and that disorder is not just an artifact of the chosen sample conditions.

3.5 Proline residues are abundant in IDPs

Further problems can occur because of the typically large proline content in the amino acid sequence of IDPs [38, 42]. As prolines do not have backbone amide protons, they are not detected in amide ^{1}H detected NMR spectra and thus represent breakpoints in the sequential backbone resonance assignment strategy based on 1HN detected triple-resonance experiments. For the same reason, variants of 2D ^{1}H-^{15}N correlation experiments typically used to follow or monitor physiological processes, such as interactions or post-translation modifications, or to measure observables such as ^{15}N relaxation rates, paramagnetic relaxation enhancements or residual dipolar couplings, do not provide information about proline residues.

3.6 IDP studies at near-physiological conditions (pH and T)

The lack of a stable 3D structure, in combination with elevated pH and temperature leads to fast exchange of amide (and other labile protein) protons with water protons. The chemical exchange rates also depend on the chemical environment (amino acid composition), and they cover a range of two orders of magnitude even for unprotected amide sites [43,44]. Fast hydrogen exchange leads to extensive line broadening of amide resonances, resulting again in missing correlation peaks and, more generally, in a heterogeneous peak intensity distribution.

4 2D HN and CON NMR experiments: the fingerprint of an IDP

Uniform ^{13}C and ^{15}N isotope labeling of proteins, and thus IDPs, in bacterial expression systems has become routinely used and it is today a necessary requirement to proceed with any high resolution NMR investigation. The nuclear spins we can deal with are thus ^{1}H, ^{13}C and ^{15}N, that means the vast majority of the nuclei in a protein.

The main considerations when designing or choosing an NMR experiment are its overall sensitivity, the resulting spectral resolution, the number of detected peaks and the information contained in the spectral parameters (peak positions and intensities). In order to optimize these parameters we can choose the source of spin polarization, the directly observed nuclei, the number and nature of the indirectly detected nuclei, the type of transfer steps used and the dimensionality of the experiment, to cite only the most important ingredients of an NMR experiment. The most useful 2D NMR experiments to characterize IDPs are introduced here.

4.1 NMR properties of ^{1}H, ^{13}C and ^{15}N in IDPs

According to equation (5), the sensitivity of the NMR experiment strongly depends on the gyromagnetic ratio of the nuclear spins (see Table 1.1) that are directly detected at the end of the pulse sequence and of those serving as polarization source at the beginning of the experiment. Because of the high gyromagnetic ratio of ^{1}H, most NMR experiments for protein applications are based on ^{1}H excitation and ^{1}H direct detection, while heteronuclear chemical shifts (^{13}C and ^{15}N) are exploited in indirect dimensions of multidimensional NMR experiments [45, 46]. However, ^{13}C detection has recently evolved into a useful tool to study biomolecules [47], in particular thanks to the larger chemical shift dispersion of ^{13}C nuclei compared to protons and to other specific benefits of the technique that improve the sensitivity of the NMR experiment, thus compensating for the lower gyromagnetic ratio, as discussed in paragraph 6.2. ^{15}N detection in protein NMR is rarely used, although some experiments have been proposed for specific applications [48–50]. NMR sensitivity is also influenced by the specific properties of the investigated system, in particular by its relaxation properties as well as by homonuclear couplings, all contributions not explicitly considered in equation (5). These aspects should also be carefully considered in the choice of the most appropriate experimental strategy to access a specific kind of information, as discussed more in detail in the next sections considering applications to IDPs.

The second important point to consider is spectral resolution. It is well known from

NMR textbooks that ^{13}C and ^{15}N nuclei show superior chemical shift dispersion compared to ^{1}H. This is definitely an interesting property provided it also holds for IDPs. As an example, Figure 3.1 shows the chemical shift dispersion observed in α-synuclein, a well-characterized IDP, for three backbone nuclear spins, $^{1}H^{N}$, ^{13}C' and ^{15}N. It is clear that also in the absence of a stable structure the chemical shift dispersion increases from $^{1}H^{N}$ to ^{13}C' to ^{15}N. The same holds true for aliphatic and aromatic spin pairs, with ^{13}C yielding a higher resolution in the NMR spectrum than the attached ^{1}H. Therefore the exploitation of the improved frequency resolution in ^{13}C and ^{15}N dimensions of multidimensional NMR spectra is crucial for the study of IDPs.

For ^{1}H and ^{13}C resonances in the side-chains of IDPs we observe a larger overall dispersion (compared to backbone nuclei) due to the differences in chemical structure between different amino acids. However, side-chain resonances from the same amino acid type cluster in the same spectral region. Side-chain resonances are therefore mainly used as indicators of the amino acid type of a given residue, while they are only of little use for the characterization of site-specific structure and dynamics in the IDP. As an example, the 2D ^{1}H-^{13}C (HC) correlation spectrum acquired on α-synuclein is shown in Figure 4.1. Despite the high sensitivity of this experiment, the overlap of signals deriving from the same type of amino acid, and the presence of resolved homonuclear couplings (see below), are responsible for the extensive overlap observed.

Since relatively narrow peaks are observed in the NMR spectra of IDPs in the absence of unfavorable dynamics, homonuclear J-couplings may contribute significantly to the linewidth. Excluding for the moment the one-bond $^{1}J_{CC}$ couplings from the discussion, which definitely need to be suppressed experimentally to obtain well-resolved ^{13}C spectra (as explained in paragraph 5.1 and 5.4), the magnitude of ^{2}J and ^{3}J homonuclear coupling constants decreases passing from ^{1}H to ^{13}C to ^{15}N, as an indirect consequence of the gyromagnetic ratios of these nuclei. This feature is thus again in favor of the detection of ^{13}C or ^{15}N instead of ^{1}H, unless the IDP has been perdeuterated to suppress homonuclear ^{1}H-^{1}H couplings.

4.2 2D ^{1}H-^{15}N and ^{13}C'-^{15}N correlation fingerprint spectra of IDPs

Recording simple 2D NMR "fingerprint" spectra of the IDP provides some useful information on the protein and allows to evaluate the overall sample properties and the feasibility of a subsequent high resolution NMR study even in the absence of a sequence-specific resonance assignment. Indeed, the number of cross-peaks detected in the spectrum compared to the number of peaks expected from the primary sequence, indicates whether the

totality of the peptide sequence is NMR-visible under the chosen sample conditions, or only part of it. The latter case occurs quite often as many proteins exhibit a significantly heterogeneous nature in terms of structural and dynamic properties leading to conformational exchange and/or aggregation induced line broadening. The observed chemical shift dispersion allows identifying whether the protein is structured, partly structured, or highly unstructured.

For such a fingerprint spectrum, we want to detect a single correlation peak per residue with good dispersion and in a reasonable amount of data acquisition time. Nuclear spins at the junction of two adjacent residues are the most suitable for this purpose. The most common experiments used are therefore ^{1}H detected ^{1}H-^{15}N (HN) [51], and ^{13}C detected ^{13}C'-^{15}N (CON) [52] correlated spectra, as shown in Figure 4.2 for human securin, a 200 amino acids IDP with more than 20% proline residues.

While the HN experiment is much more sensitive and can be recorded in a significantly shorter time, proline residues are only detected in the CON spectrum that also shows a better spectral resolution. In addition, the CON spectrum does not suffer from hydrogen-exchange-induced line broadening and thus can still be recorded under conditions of high pH and temperature, where many HN peaks are no longer detectable [6], as shown in Figure 4.3 that reports the HN and CON spectra acquired on α-synuclein with increasing temperature, approaching physiological conditions. On the other hand, high-quality HN correlation spectra can be recorded on protein samples at concentrations of only a few μM on a spectrometer equipped with a cryogenic probe. The 2D HN and CON spectra are thus highly complementary both in terms of detectability and of information content. These 2D correlation experiments can be acquired in different ways. The most appropriate variants for applications to IDPs are discussed in detail in the following section.

These 2D spectra can also be used to follow changes in the properties of the IDP upon changing the experimental conditions such as temperature, pH, ionic strength, buffer, reducing/oxidizing environment, or additions of potential partners such as metal ions, small molecules, nucleic acid fragments and proteins. They also enable to follow chemical reactions such as the occurrence of post-translational modifications [53]. Finally they provide an invaluable tool to take a snapshot of a protein inside an entire cell [54–56]. As an example, the 2D HN and CON experiments acquired on α-synuclein in cell are compared with those acquired on the purified protein (Figure 4.4), showing that they can be used for the investigation of IDPs in-cell.

Finally, modifications of these basic 2D experiments (HN and CON) enable to determine a variety of observables that report on different properties of the IDP at atomic

resolution, once sequence-specific assignment becomes available. These include ^{15}N relaxation rates, scalar couplings, residual dipolar couplings (RDCs), paramagnetic relaxation enhancements (PREs), cross-relaxation (σ_{HH}) and cross-correlation rates (CCR), as well as solvent exchange rates. All these NMR observables report on the structural and dynamic features of the IDP and will be discussed in detail in chapter by Konrat.

5 NMR tools for overcoming major bottlenecks of IDP studies

5.1 Multidimensional NMR, indirect frequency editing and non-uniform sampling

The large number of NMR-active nuclei in a protein results in severe resonance overlap in one-dimensional spectra (^{1}H, ^{13}C, ^{15}N), making it practically impossible to extract atom-resolved information from them. This problem is circumvented by using multidimensional (nD) NMR techniques that spread and correlate the signals of individual nuclear spins along different frequency dimensions [57]. Multidimensional NMR data are recorded by repeating the basic pulse sequence numerous times. For each repetition of the experiment (transient), the observable NMR signal is detected along one dimension, while additional 'indirect' dimensions are sampled by incrementing a time variable in the pulse sequence from one repetition to the next. Despite the great success of multidimensional NMR, an important drawback of this stepwise sampling procedure is the long experimental time, a direct result of the hundreds or even thousands of transients that are required for a single data set (Figure 5.1).

In this paragraph we will discuss different editing techniques to improve the spectral resolution in the indirect dimensions of nD NMR experiments, and advanced non-uniform sampling (NUS) approaches that are indispensable to record high-dimensional spectra in reasonable experimental times.

5.1.1 Real-time, constant-time and semi-constant time frequency editing

Frequency editing of nuclear spins is the key step to introduce indirect dimensions in multidimensional experiments. In order to do so, a pulse sequence element that allows time evolution of the frequency-edited spins (I), while suppressing (refocusing) all other chemical shift and J-coupling evolutions, is required. Figure 5.2 shows five common pulse sequence implementations of I-spin editing during an incremented time variable t_1: (a) conventional real-time, (b) optimized real-time, (c) and (d) constant-time (CT), and (e) semi-CT editing.

In the conventional real-time implementation, heteronuclear J_{IS} and J_{IK} coupling evolutions are refocused by a 180° pulse applied in the middle of t_1. In this context "heteronuclear" means that the particular spin species can be manipulated separately by an appropriately shaped rf. pulse. As an example, ^{13}C' and ^{13}C$^{\alpha}$ can be manipulated

separately, as they have well-separated chemical shift ranges, while $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ have generally overlapping chemical shift ranges, and thus cannot be separated spectroscopically. Therefore, the $^{13}C^{\alpha}$-$^{13}C^{\beta}$ coupling gives rise to a peak splitting in the $^{13}C^{\alpha}$ dimension of a NMR spectrum recorded with conventional real-time editing as illustrated in Figure 5.3a. A slightly optimized version of real-time editing is shown in Figure 5.2b, where chemical shift evolution of the I-spins during the 180° S-spin decoupling pulses, as well as Bloch-Siegert phase shifts, are refocused. This allows recording a data set that does not require any phase correction in the corresponding indirect dimension (t_1), in order to avoid baseline distortions.

In order to avoid line splitting due to homonuclear J_{II}-coupling evolution, the CT editing blocks of Figure 5.2c and 5.2d can be used, with the CT delay T set to $T = 1/J_{\mathrm{II}}$. A $^{1}H^{N}$-$^{13}C^{\alpha}$ correlation spectrum recorded with the pulse sequence block of Figure 5.2c is shown in Figure 5.3b. The detected NMR signal in this spectrum is modulated by a factor $\cos^{n}(\pi J_{CC}T)$, with n is the number of carbon atoms attached. In addition, signals with different sign provide meaningful information about the spin-coupling topology, as for example correlation peaks of glycine residues are of opposite sign with respect to all others, as glycines have no $^{13}C^{\beta}$ attached to the $^{13}C^{\alpha}$. Alternatively, CT frequency editing can be applied during an INEPT-type coherence transfer step [58]. This is shown in Figure 5.2d where the I-spins are edited during an I to S transfer. Here, the CT delay has to be set to $T = 1/2J_{\mathrm{IS}}$ in order to achieve complete transfer.

A drawback of CT editing is that the maximum possible evolution time (t_1^{max}) is limited by the CT delay T so that $t_1^{max} \leq T$. This, of course, has an effect on the achievable spectral resolution that is a crucial point for NMR studies of IDPs as discussed before. A possible solution could be to increase the CT delay to $n \times T$. However also introducing too long delays in the pulse sequences may cause pronounced relaxation losses. Therefore semi-CT editing has been proposed to enhance the spectral resolution while still exploiting spin evolution during a coherence transfer delay. The improvement achieved with semi-CT evolution in terms of spectral resolution in the ^{15}N dimension of an HNCA experiment is illustrated in Figure 5.4.

5.1.2 Strategies for non-uniform data sampling in multi-D NMR

The amplitude of the NMR signal, or free induction decay (FID), is typically measured (sampled) at discrete, uniformly spaced time points. In a multidimensional NMR experiment, the same approach is generally employed also in the indirect dimensions, since uniform sampling is required for data processing using the fast Fourier transform (FFT).

The Nyquist theorem states that the sampling rate needs to be faster than twice the highest frequency expected or, in other words, determines a maximum value for the time interval between sampled points (dwell time) that has to be used to avoid spectral folding of the peaks. Hence, the sampled data points form a Cartesian grid with the spacing between time points in each dimension given by the inverse of the spectral width of the edited nuclei. As a result of this uniform data sampling procedure, the experimental time requirement increases roughly by about two orders of magnitude for each additional dimension, making high-dimensional experiments (> 3D) impractical or requiring strong compromises on the spectral resolution in the indirect dimensions.

In order to overcome these limitations, alternative sampling strategies combined with appropriate processing tools have been developed over the past two decades. These sparse or non-uniform sampling (NUS) techniques rely on a reduction of the overall number of sampled time points by recording only a subset of the data points of the Cartesian grid. Examples of alternative sampling schemes are shown in Figure 5.5.

Some of these sampling grids, e.g. linear under sampling or radial sampling, still yield data sets that can be processed using FFT [59–61]. The general NUS scheme, however, where a certain percentage of the grid points is randomly chosen, requires alternative processing tools. Currently, several algorithms for processing non-uniformly sampled data are available and well-established, such as multidimensional decomposition (MDD) [62,63], maximum entropy (MaxEnt) methods [64], compressed sensing (CS) [65,66], multidimensional Fourier transform (MFT) [67] and spectroscopy by integration of frequency and time domain information (SIFT) [68]. The common feature of all these methods is that they aim in finding the NMR spectrum that, when applying inverse Fourier transform, best reproduces the measured time data points. As in the case of sparse data sampling, this is an underdetermined computational problem. Therefore, some additional assumptions have to be made to choose the most likely spectrum out of the manifold of possible solutions. For example, MDD exploits the prior knowledge that NMR signals are the direct product of Lorentzian (or Gaussian) line shapes in each frequency dimension. MaxEnt relies on the optimization of a penalty function in order to select the spectrum that has the highest entropy (minimal number of signals). Similarly, CS performs a $l1$-norm minimization to recover the sparsest spectrum, while MFT performs discrete multidimensional Fourier transform of the data using appropriate data weighting and filtering for the reduction of sampling noise. Finally, SIFT is an iterative method that replaces the missing time domain data points by zero for the first iteration. This allows classical FFT processing. In subsequent iterations, the missing data points are replaced by the result of inverse FT

of the obtained NMR spectrum after having set to zero known empty spectral regions.

At present, the question which is the most appropriate sampling grid and processing technique for a given experiment is non-trivial, but a few general recommendations can be given: (i) the percentage of sampled data points can be decreased with increasing dimensionality of the experiment and sparseness of the final NMR spectrum (number of expected correlation peaks). Typically, in the case of a 4D backbone assignment experiment a few percent of sampled data points are sufficient. (ii) A random distribution of the sampled data points is particularly recommended to spread the sampling noise over the entire frequency domain (with apparent reduction of its overall amount), preventing artifacts to be clustered in specific spectral regions. (iii) The choice of the processing algorithm mainly depends on the available software (degree of automation, required expertise for proper parameter adjustment) and the computer power (some algorithms require more computer power and memory than others). If possible, it is always a good option to test different processing tools on the same data set. Recent developments by spectrometer manufacturers make it straightforward today to set up a random sampling scheme and to run experiments in NUS mode. Also some of the non-linear processing routines, such as MDD and CS, have been interfaced with the NMR spectrometer software, and are available for routine use. Alternatively, most of the processing algorithms can be accessed online and downloaded. These advanced data acquisition and processing tools should be largely exploited for NMR investigations of IDPs, as they allow performing experiments of high dimensionality in a reasonable amount of time at very high spectral resolution as discussed in detail in section 7.

5.2 Selecting ^{15}N spin states with favorable transverse relaxation properties

Multi-dimensional NMR correlation experiments require a series of coherence transfer steps and frequency editing periods, during which the detectable NMR signal, and thus the sensitivity of the experiment, decreases due to transverse spin relaxation. In order to improve the performance of such experiments and therefore their applicability to large IDPs and low sample concentrations, it is important to select the coherences with the longest transverse relaxation times for the transfer and chemical shift editing steps. Here we present two commonly used techniques for reducing signal losses during ^{15}N transverse evolution periods as required for all amide ^{1}H detected NMR experiments, and that are also of importance in ^{13}C detected experiments.

In a standard INEPT-type ^{15}N$\rightarrow$ ^{13}C transfer step, ^{15}N single quantum coherence

(SQC) oscillates forth and back between in-phase (N_x) and anti-phase ($2N_yH_z$) coherence, as a result of ^{15}N-^{1}H scalar coupling evolution. As a consequence, the effective relaxation during the transfer is given by the average of in-phase and anti-phase SQ relaxation rates. In uniformly ^{13}C, ^{15}N labelled proteins, relaxation of anti-phase SQC is about 30% faster than relaxation of in-phase SQC. This difference is even more pronounced in the case of significant solvent exchange rates between the amide and water protons, further reducing the lifetime of anti-phase SQC. Improved transfer efficiency is thus achieved by ^{1}H composite pulse decoupling that removes the ^{15}N-^{1}Hscalar coupling evolution, and maintains the in-phase SQC throughout the duration of the transfer. Common ^{1}H decoupling schemes are MLEV-16 [69], DIPSI-2 [70], WALTZ-16 [71,72] and GARP-1 [73]. They are all composed of a basic pulse sequence element (R) performing basically a broadband 180° spin inversion, and that is repeated by applying an additional phase cycle. They differ mainly by the B_1 field strength required to achieve a certain decoupling band width.

A different approach to enhance spectral resolution in ^{15}N spectra, especially at high magnetic field strength, is transverse relaxation-optimized spectroscopy (TROSY) introduced by Pervushin et al. in 1997 [74]. The TROSY effect is based on the interference between two different spin relaxation mechanisms, e.g. the dipolar coupling (DD) and the chemical shift anisotropy (CSA), in a scalar coupled spin pair such as ^{1}H-^{15}N. This interference can be constructive (increasing the apparent relaxation rate) or destructive (decreasing the apparent relaxation rate) for different components of the peak multiplet (see Figure 5.6). The TROSY pulse sequence allows the selection of the mutliplet component (single-transition spin states) giving the sharpest lines [74]. The TROSY effect for ^{1}H-^{15}N is highest at magnetic field strengths of about 1 GHz, and increases with the effective tumbling correlation time. TROSY-based pulse sequences are thus especially useful to enhance peak intensities and line shapes of IDP regions that are involved in transient structure formation and/or at low temperatures.

5.3 Longitudinal-relaxation enhancement for increased sensitivity and reduced acquisition times

In this paragraph, we discuss the dependence of the experimental sensitivity on the recovery delay T_{rec} [75]. A schematic drawing of an NMR experiment is shown in Figure 5.7, consisting in a pulse sequence of length t_{seq}, a data acquisition period t_{acq}, and an additional inter-scan delay t_{rec}. A recovery delay is required for relaxation of the system to restore sufficient spin polarization to restart the experiment for a subsequent scan ($T_{rec} = t_{acq} + t_{rec}$). Its duration depends on the longitudinal relaxation time constant T_1.

On one hand, the detected signal intensity is proportional to the amount of spin polarization available at the beginning of each scan under steady-state conditions. On the other hand, the SNR scales with the square root of the number of scans that can be performed during a given experimental time, and thus the scan time $T_{scan} = t_{seq} + t_{acq} + t_{rec}$. These two effects can be described analytically by the following equation:

$$SNR \propto \frac{1}{(1 - \exp(-T_{rec}/T_1))\sqrt{T_{scan}}} \tag{9}$$

Equation (9) implies that maximum sensitivity is obtained by adjusting the recycle delay T_{rec} to:

$$T_{rec}(SNR_{max}) \widetilde{=} 1.25 T_1 \tag{10}$$

and thus the dependency of the experimental sensitivity on the longitudinal relaxation time T_1 is approximately given by:

$$SNR_{max} \propto \frac{1}{\sqrt{T_1}}. \tag{11}$$

Enhancing the longitudinal relaxation efficiency of the excited spins thus provides an interesting way to increase the experimental sensitivity and additionally to reduce the minimal data acquisition time. Here we will focus on ^{1}H excitation experiments. In order to understand the experimental schemes that have been proposed for longitudinal proton relaxation enhancement, we need to briefly discuss the spin interactions that govern proton relaxation. There are essentially two different relaxation mechanisms that are responsible for proton longitudinal relaxation in a protein: (i) ^{1}H-^{1}H dipolar interactions; (ii) hydrogen exchange processes between labile protein protons, e.g. amide and hydroxyl protons, and water solvent protons. The time evolution of the polarization of each proton spin in the molecule is given by the Solomon or Bloch-McConnell equations [2, 76, 77]:

$$-\frac{d}{dt}\begin{pmatrix} W_{1z}-W_{1z}^0 \\ H_{1z}-H_{1z}^0 \\ H_{2z}-H_{2z}^0 \\ \vdots \\ H_{nz}-H_{nz}^0 \end{pmatrix} = \begin{pmatrix} \rho_W & 0 & 0 & \cdots & 0 \\ 0 & \Sigma_j\rho_{1j} & \sigma_{12} & \cdots & \sigma_{1n} \\ 0 & \sigma_{21} & \Sigma_j\rho_{2j} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \sigma_{n1} & \sigma_{n2} & \cdots & \Sigma_j\rho_{1nj} \end{pmatrix} \begin{pmatrix} W_{1z}-W_{1z}^0 \\ H_{1z}-H_{1z}^0 \\ H_{2z}-H_{2z}^0 \\ \vdots \\ H_{nz}-H_{nz}^0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ -k_{ex,1} & k_{ex,1} & 0 & \cdots & 0 \\ -k_{ex,2} & 0 & k_{ex,2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -k_{ex,n} & 0 & 0 & \cdots & k_{ex,n} \end{pmatrix} \begin{pmatrix} W_{1z} \\ H_{1z} \\ H_{2z} \\ \vdots \\ H_{nz} \end{pmatrix} \quad (12)$$

where H_{iz} denotes the z-component of the polarization of proton i and H_{iz}^0 is its thermal equilibrium value. The different ρ and σ terms stand for auto- and cross-relaxation rate constants, with values depending on the distance separating the two protons involved as well as the global and local dynamics of the protein experienced at the sites of the two protons. W_z stands for the bulk water polarization, and $k_{ex,i}$ are the hydrogen exchange rates for individual protons with the water.

Equation (12) indicates that the relaxation of an individual proton spin depends on the spin state of all other protons in the protein, as well as the bulk water, at the start of the recovery time. The selective spin manipulation of a subset of protons, e.g. amides, while leaving all other protein and water proton spins unperturbed, thus provides an efficient spectroscopic tool for enhancing longitudinal proton spin relaxation, as will be shown in the following paragraph.

The contributions of solvent exchange and dipolar cross-relaxation of amide protons with unperturbed aliphatic proton spins depend on the size and residual structure of the IDP, and the sample conditions. Figure 5.8 shows apparent amide ^{1}H T_1 values measured for different IDP samples upon selective or non-selective inversion of different sets of proton spins [6,78]. For NS5A and BASP1, studied at low temperature and pH, the major relaxation enhancement mechanisms are dipolar interactions, while this situation changes as solvent exchange becomes more efficient, as shown for α-synuclein and PV core protein that were studied at higher temperature and pH. The average T_1 values measured for the different IDPs shown in Figure 5.8, as well as the range of predicted exchange rates using the SPHERE program are given in Table 5.1 [43, 44].

When dipolar interactions provide the dominant relaxation mechanism, non-selective

^{1}H T_1 values are of the order of 900 ms, while selective spin manipulation allows to reduce them by a factor of three to four, reaching values of 200-300 ms [46, 79]. This difference becomes even more pronounced in the case of fast hydrogen exchange. Under these conditions, the non-selective ^{1}H T_1 approach the T_1 of bulk water ($\sim$ 3 s at 25 °C), while the selective T_1 become as short as 60 ms [6]. For α-synuclein at pH 7.4 and 15 °C, the average amide ^{1}H T_1 is reduced by a factor of 38, resulting in a potential sensitivity gain of a factor of six, according to equation (11). This clearly motivates the use of amide ^{1}H start pulse schemes that leave aliphatic and water protons unperturbed throughout the NMR experiment.

The simplest, and probably most efficient way to achieve longitudinal relaxation enhancement (LRE) is the use of band-selective ^{1}H pulses, especially in the case of amide protons resonating in a frequency range that is well-separated from aliphatic and water protons. Pulse sequences exploiting this concept have been termed Band-selective Excitation Short-Transient (BEST) experiments [46, 51, 78, 80]. BEST pulse sequence elements are available for all basic pulse sequence elements (building blocks) [29], e.g. INEPT, sensitivity-enhanced reverse-INEPT (SE-RINEPT or planar mixing) and TROSY (or double-S3CT) required for setting up most of the common triple-resonance experiments. In this paragraph we will focus on BEST-TROSY pulse sequences that have been proven to be particularly useful for the study of IDPs using high magnetic field NMR instruments.

BEST-TROSY combines the advantages of longitudinal relaxation optimization, discussed in the previous paragraph, with those of transverse relaxation-optimized spectroscopy, discussed in paragraph 5.2. A special feature of BEST-TROSY is that proton (H_z) polarization that builds up during the pulse sequence as a consequence of spin relaxation is converted to ^{15}N polarization (N_z) by the final ST2-PT sequence element [51]. In a conventional TROSY sequence, a large portion of this additional polarization is lost due to ^{15}N T_1 relaxation during the long recovery delay that restores ^{1}H polarization. In BEST-type experiments, however, most of the polarization is conserved because of the short recycle delays used. This provides an additional mechanism for sensitivity and resolution improvement.

The ^{1}H-^{15}N BEST-TROSY sequence and spectra recorded for two large (250 residue) IDPs are shown in Figure 5.9. Long t_1 acquisition times in the ^{15}N dimension result in highly resolved 2D correlation maps, despite the low chemical shift resolution observed in the ^{1}H dimension. The high resolution obtained in the ^{15}N dimension of a 2D BEST-TROSY (BT) spectrum can also be transferred to higher-dimensional BT ^{1}H-^{13}C-^{15}N or

BT ^{1}H-^{15}N-^{15}N correlation experiments, as required for sequential resonance assignment (paragraph 6.2), by using the semi-CT editing scheme discussed in paragraph 5.1. Furthermore, it has been shown that BEST-TROSY implementations of such experiments provide a 20% to 80% increased SNR compared to BEST-HSQC versions, for different IDP samples at 800 MHz ^{1}H frequency [78]. BT also results in a more uniform distribution of peak intensities in the spectrum, as the signal enhancement is efficient for residues in transiently structured peptide regions which tend to be the ones characterized by smallest intensities in optimal conditions for 2D HN correlation experiments.

5.4 ^{13}C detected experiments

^{13}C direct detection, routinely used to study small molecules, has recently become a useful tool also for biomolecular applications [81] thanks to the impressive improvements in instrumental sensitivity [4] and to the development of new NMR experiments. The key improvements that have been important to convert ^{13}C direct detection into a tool useful also for uniformly labeled protein samples, as well as the important aspects that have to be considered for a proper selection of the most appropriate experimental variant to focus on a specific problem, are discussed hereafter.

The first aspect to consider when performing ^{13}C direct detection on uniformly labeled protein samples consists in the presence of large splittings of the ^{13}C resonances in the direct acquisition dimension due to the presence of large homonuclear one-bond couplings, a feature that definitely represents a novelty with respect to ^{1}H direct detection. Indeed, despite ^{13}C-^{13}C couplings are very useful in the design of multidimensional NMR experiments, they are responsible for large signals splitting that, as evident from Figure 5.10, need to be suppressed in order to preserve high resolution in the direct acquisition dimension.

5.4.1 Homonuclear ^{13}C decoupling

Achieving homonuclear ^{13}C decoupling in the direct dimension is more complex than in the indirect dimensions, because it typically requires the application of radiofrequency pulses at a similar frequency than those of the nuclear spins that are in the process of being detected. A possible solution to this problem consists in sharing the detection period between data acquisition and decoupling pulses (band-selective homonuclear decoupling). This, however, reduces the overall sensitivity of the experiment, and introduces Bloch-Siegert phase shifts [82] and decoupling side bands [83] in the spectrum. Recently improved

variants exploiting this general idea that however use hard pulses applied at regular intervals during the acquisition period have been proposed [84]. Alternatively, post-acquisition methods, such as data deconvolution using maximum entropy reconstruction, may be used [85]. A last, and arguably the most elegant class of methods for ^{13}C homonuclear decoupling [86], also known as "virtual decoupling", uses spin-state selection.

Virtual J_{CC} decoupling uses an additional spin evolution delay prior to detection, and requires the recording of, at least, two experiments with different parameter settings. In the first experiment, spin evolution under the J_{CC} coupling is suppressed resulting in an in-phase (IP) line splitting in the detected spectrum, while in the second experiment the J_{CC} coupling evolves for a time $1/(2J_{CC})$ resulting in anti-phase (AP) line splitting in the final spectrum. A single resonance line (without splitting) is then obtained by calculating the sum and the difference of the two recorded spectra. Finally, the two resulting (sum and difference) spectra are shifted by the amount $J_{CC}/2$ with respect to each other and added up to yield a single line at the correct chemical shift position [52, 87–89]. This approach, illustrated in Figure 5.11 for virtual decoupling of $^{13}C^{\alpha}$ in the NMR spectrum of carbonyls, is at the basis of most of the ^{13}C detected multidimensional experiments. In order to work properly, virtual J_{CC} decoupling requires quite uniform J_{CC} couplings in the protein which is the case for the large one-bond coupling between alpha carbons ($^{13}C^{\alpha}$) and carbonyls ($^{1}J_{C\alpha C'} \approx 53$Hz). It is worth noting that the two experiments rely on the same number of pulses and lengths of delays, in order to ensure identical signal loss due to pulse imperfections and spin relaxation effects. Another prerequisite of this technique is that the two nuclear spins, e. g. ^{13}C' and $^{13}C^{\alpha}$ are sufficiently well separated to allow their selective manipulation through band-selective pulses. There is also a price to pay in terms of sensitivity for the virtual decoupling method, as during the linear combinations necessary to achieve homonuclear decoupling also the thermal noise is increased, thus leading to a reduction of the signal to noise ratio.

Virtual decoupling can be appended to any pulse sequence ending with in-phase ^{13}C' transverse coherence [90]. Most of the triple resonance experiments based on ^{13}C' direct detection end with a coherence transfer step that involves refocusing of either ^{13}C'-^{15}N or ^{13}C'-$^{13}C^{\alpha}$ antiphase coherence (or both). Therefore, virtual decoupling can be implemented without appending an additional block by slightly changing the position of the $^{13}C^{\alpha}$ inversion pulses in the last coherence transfer steps, as illustrated in Figure 5.11.

Finally, as these spin-state selective approaches to achieve homonuclear decoupling perform so well, it is worth mentioning that they can also be implemented for heteronuclear decoupling [91, 92]. Indeed, a clever variant to achieve heteronuclear ^{15}N decoupling has

been proposed both for $^{1}H^{N}$ and $^{13}C'$ direct detection experiments. This can be useful when the $^{13}C'$ coherence lifetimes allow for long acquisition times or to reduce the radiofrequency load/heating on the ^{15}N channel during acquisition, which results particularly useful in the fast pulsing regime [6].

5.4.2 Starting polarization source

The other important point to consider in the design/choice of a ^{13}C detectedexperiment is the starting polarization source. From paragraph 4.1, it is immediately clear that a relevant contribution to increase the sensitivity of ^{13}C detected experiments can come from the use of ^{1}H as starting polarization source [92–94]. In fact, the large one-bond scalar couplings can be easily used to transfer polarization from ^{1}H to the directly bound heteronuclei while still keeping experiments "exclusively heteronuclear". This means that only heteronuclear chemical shifts are frequency labeled in all the spectral dimensions to take advantage of their larger chemical shift dispersion compared to that of ^{1}H, but still exploiting the larger ^{1}H polarization deriving from its higher gyromagnetic ratio. Therefore, several exclusively heteronuclear NMR experiments, such as the (H)CBCACON and (H)CBCANCO experiments [92], feature ^{1}H as a starting polarization source.

With protons exploited only to increase the experimental sensitivity, it is worth thinking about the possibility to implement ^{1}H longitudinal relaxation enhancement techniques to reduce the duration of NMR experiments or to increase the sensitivity per unit of time or the resolution. Actually, in these experiments it is very easy to manipulate different sets of ^{1}H spins in a selective way, an essential requirement to promote LRE. Indeed, a simple modification of the initial INEPT block can be introduced to flip-back all the other spins not directly involved in the coherence transfer pathway (H^{flip} method [95, 96]).

The solution described above for amide protons, based on the use of band-selective pulses on amide protons (BEST approach), can also be implemented in ^{13}C detected experiments [6, 97]. Amide protons indeed resonate in a quite isolated region of the ^{1}H spectra, well separated from the water resonance, and therefore they can be selectively manipulated leaving both water and aliphatic resonances unperturbed.

As discussed above, the LRE effect deriving from exchange processes with the solvent is very pronounced in IDPs and it adds to that deriving from ^{1}H-^{1}H NOEs, which may contribute at low temperature conditions or in the presence of partially structured elements, even if with a minor extent. Approaching physiological conditions (high temperature and neutral pH), the LRE effect is so large that the recovery of amide polarization to equilibrium becomes extremely fast and recycle delays between transients are therefore not

necessary anymore. In fact, in this type of experiments the longitudinal recovery already starts before the end of the pulse sequence, as protons get perturbed only in the very initial part of the experiment. Therefore, the optimal set up consists in an inter-scan delay equal to zero as implemented for the $H^{N\text{-}BEST}CON$ experiment [6]. Its high sensitivity enables to collect 2D spectra in a couple of minutes, as shown in Figure 5.12, making it an ideal technique for *in-cell* applications. However, the use of $^1H^N$ as starting polarization source renders the experiment susceptible to the loss of information about proline residues and reintroduces a dependence on exchange rates of amide protons with the solvent that impacts on the experimental sensitivity.

Alternatively $^1H^\alpha$ as starting polarization source can be exploited. In this way, all backbone sites can be sampled (including prolines) and exchange processes with the solvent are avoided. On the other side of the coin, this also means that only LRE effects deriving from ^{1}H-^{1}H NOEs can be exploited. Therefore, LRE effects become sizeable either at lower temperature or for partially structured proteins. The most convenient way of achieving LRE for $^1H^\alpha$ consists in exploiting the one-bond scalar coupling ($^1J_{C\alpha H\alpha}$) with the attached carbon ($^{13}C^\alpha$) to selectively manipulate $^1H^\alpha$ and flip-back all other protons that are not actively used in the magnetization transfer pathway (H^{flip} method) [95]. Therefore, this trick can be easily implemented in any $^1H^\alpha$-start pulse sequence to achieve LRE. Also increasing the temperature provides a simple tool to increase longitudinal relaxation of $^1H^\alpha$ protons that does not require selective manipulation of $^1H^\alpha$ protons.

Finally, ^{13}C-start versions of such ^{13}C detected experiments can be designed. Indeed ^{13}C spins are not directly involved in chemical exchange processes and are characterized by large chemical shift dispersion. The high flexibility of IDPs results in relatively short ^{13}C T_1 values, which are on the order of second, so that recovery delays remain sufficiently short. In addition, the ^{1}H-^{13}C NOE effect can be exploited in the case of highly flexible IDPs to enhance the signal-to-noise ratio just by irradiating protons during the recovery delay [96]. Therefore, in many cases of practical interest the sensitivity of ^{13}C-start-^{13}C detected versions of experiments, in particular for the simplest 2D experiments such as the CON, CACO and CBCACO, is sufficient to obtain informative spectra [98].

6 From 2D to 3D: from simple snapshots to site-resolved characterization of IDPs

Multidimensional NMR experiments that encode chemical shift information in several indirect dimensions provide the necessary resolving power to separate correlation peaks from different sites in an IDP. In this paragraph, the general strategy used to achieve site-specific resonance assignment of the NMR signals will be discussed, and a selection of commonly used multidimensional experiments for sequential resonance assignment, as well as complementary techniques for amino acid type identification will be presented.

6.1 Sequential NMR assignment: the general strategy

In order to extract site-specific information from the NMR experiments, each resonance observed in the spectra needs to be associated to a specific nuclear spin of the protein. This task is commonly called *sequence-specific* (or *sequential*) *resonance assignment*. In the early days of biomolecular NMR spectroscopy, resonance assignment was based on the information contained in 2D homonuclear ^{1}H correlation experiments [99]. The combined analysis of a set of through-bond (COSY, TOCSY) and through-space (NOESY, ROESY) spectra allows sequence-specific assignment of well-folded proteins of moderate size ($\sim$100 residues). In the case of highly disordered proteins and IDPs, characterized by low chemical shift dispersion in the 2D ^{1}H spectra, this strategy is limited to peptides of less than a few tens of residues. As a consequence, NMR studies of IDPs require uniform ^{13}C and ^{15}N labeled samples that enable the use of 3D (or higher-dimensional) through-bond-only experiments for sequential resonance assignment purposes [100, 101]. These experiments are based on a series of coherence transfer steps, which exploit the large one- and two-bond scalar couplings (Figure 6.1) [29, 45].

Starting from a chosen polarization source, typically amide ^{1}H, aliphatic ^{1}H, or ^{13}C, a series of subsequent coherence transfer pathway steps allows correlating NMR frequencies of different nuclear spins in the protein with high efficiency. Ideally, this strategy could be used to transfer coherence (and thus information) all along the polypeptide chain. In practice, due to spin relaxation effects, coherence transfers are generally limited only to nuclear spins within the same or neighboring residues. Therefore, a set of spectra providing complementary connectivity information is required for sequential resonance assignment. The most widely used experiments are introduced in the next paragraphs. They can be divided into 3 categories: (i) intra-residue, (ii) sequential, and (iii) bi-directional correlation experiments, according to the type of information they contain. A schematic drawing

showing how a pair of such spectra can be used in the sequential resonance assignment procedure is reported in Figure 6.2.

The basic strategy for sequential assignment is that the different nuclei correlated in a 2D HN or 2D CON fingerprint spectrum, as introduced in paragraph 4.2, are connected in a 3D data set with one (or more) additional nuclei, e.g. $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{13}C'$, or $^{1}H^{\alpha}$, that are frequency labeled in the third dimension (indicated as X nuclear spins). ^{1}H-^{15}N or $^{13}C'$-^{15}N pairs are then recognized as belonging to neighboring residues when the corresponding X frequencies match. In case of identical or similar resonance frequencies of two or more X nuclei, ambiguities can be solved by combining the information obtained from different pairs of intra- and inter-residue 3D H-N-X or CO-N-X spectra. As a result of this assignment step, chains of sequentially connected residues (^{1}H-^{15}N or $^{13}C'$-^{15}N pairs) are identified. This so-called "sequential assignment walk" might be interrupted by missing correlation peaks, either due to unobservable (line-broadened) NMR signals, e.g. because of pronounced conformational exchange processes or chemical exchange at physiological conditions, missing correlation information, e.g. amide ^{1}H in proline residues, and/or remaining frequency ambiguities; the combination of 3D H-N-X and 3D CO-N-X spectra provides more complete information reducing interruptions in the sequence-specific assignment, ambiguities contributing in this way to the completeness and accuracy of assigned resonances. The final assignment step consists in mapping the identified peptide fragments (correlated NMR nuclei) to their specific positions in the protein sequence. In order to do so, amino acid type information on some (or all) of the residues that form the fragment is required, as illustrated in Figure 6.3. Such amino acid type information is obtained from side-chain ^{13}C chemical shifts, or from specifically tailored experiments that use spectral editing techniques to differentiate spin-coupling topologies in the amino acid side-chains (see paragraph 6.3).

6.2 Sequential NMR assignment: 3D experiments

6.2.1 $^{1}H^{N}$ detected experiments

The most common and established resonance assignment approach for proteins uses a set of $^{1}H^{N}$-detected 3D correlation experiments [102, 103]. The resulting spectra share as common frequencies those of amide protons in the direct dimension and of amide nitrogen atoms in one of the indirect dimensions. The most useful experiments and their underlying coherence transfer pathway(s) are briefly presented in this section (Figure 6.4). As discussed in paragraph 5.3, most amide ^{1}H detected experiments for IDP studies should

be implemented as BEST or BEST-TROSY versions in order to enhance experimental sensitivity and spectral resolution [46, 78].

The most sensitive coherence transfer pathway correlates the amide $^{1}H^{N}$ and ^{15}N of amino acid i with the $^{13}C'$ of the preceding $(i-1)$ residue. This experiment, called 3D HNCO [78, 103–105] is often used to "count" the number of cross-peaks and so to estimate the number of resolved spin systems (residues) that are observed. Additional transfer steps allow correlating the amide $^{1}H^{N}$ and ^{15}N with the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ of the preceding $(i-1)$ residue, in the so-called 3D HN(CO)CA [78, 104, 106] and 3D HN(CO)CACB [78, 107, 108] experiments. Note that the intrinsic sensitivity is decreasing with each additional transfer step, making HN(CO)CACB the least sensitive experiment of the series. For sequential resonance assignment, the information obtained from the experiments correlating nuclei of neighboring amino acids needs to be complemented with that provided by experiments correlating the amide $^{1}H^{N}$ and ^{15}N with the $^{13}C'$, $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ of the same (i) residue. This can be achieved, for example, by acquiring 3D HNCA [46, 103, 104], 3D HNCACB [46, 109, 110] and 3D HN(CA)CO [46, 111–113] experiments, belonging to the class of "bi-directional experiments". All of these experiments result in two cross-peaks per residue, because of the similar size of $^{1}J_{NC\alpha}$ and $^{2}J_{NC\alpha}$ coupling constants (Figure 6.1), one corresponding to the desired intra-residue correlation, the other one to the sequential correlation already detected in the previous set of experiments. In order to avoid recording of redundant information and to limit the risk of peak overlaps in the spectra, purely intra-residue correlation experiments, 3D iHNCA [78, 114, 115], 3D iHNCACB [78, 114, 115] and 3D iHN(CA)CO [78, 116] have been proposed as an alternative to the bi-directional experiments discussed above. The bi-directional and intra-residue experiments are less sensitive than their sequential counterparts and thus, as a rule of thumb, the number of scans should be at least doubled. Recently, improved HNCA+, HNCACB+, and HNCO+ pulse sequences have been introduced to perform bi-directional experiments with improved sensitivity for the sequential correlations [117]. These experiments are especially useful for the study of fragile IDP samples as they allow retrieving the complete information required for sequential assignment from a single set of 3D experiments.

An additional class of experiments, especially useful for IDP samples, is referred to as 3D H-N-N, since their main characteristic is that ^{15}N frequencies are labeled in both indirect dimensions. This allows performing the sequential assignment walk by matching ^{15}N frequencies that, as discussed in paragraph 4.1, experience the largest chemical shift dispersion for IDPs lacking a stable structure. A drawback of these pulse schemes is that they require more or longer transfer steps, thus resulting in reduced sensitivity especially

for IDP residues involved in transient secondary structure formation. The coherence transfer pathways for the 3D (H)N(COCA)NH and 3D (H)N(CA)NH experiments [118–122] are shown in Figure 6.5. In the 3D (H)N(COCA)NH, the HNCO sequence is extended passing coherence from ^{13}C' to $^{13}C^{\alpha}$ and then to ^{15}N of the neighboring amino acid, resulting in the correlation of $^{1}H^{N}$ and ^{15}N of amino acid i with the ^{15}N of amino acid i and $i+1$, while the 3D (H)N(CA)NH is an extension of HNCA, yielding correlations with ^{15}N of residues $i-1$, i, and $i+1$. In the case of highly flexible IDPs resulting in large T_2 values, these pulse schemes can be optimized to detect mainly the sequential correlations, while suppressing to a large extent the "diagonal" N_i-N_i correlation peak [118].

A last class of experiments allows to extend the assignment to the side chain ^{13}C resonances, and to provide valuable information on the side chain length and ^{13}C chemical shifts, important for distinguishing between amino acid types. The most widely used pulse sequence is called 3D (H)C(CO)NH-TOCSY [123–127]. The sequence starts from aliphatic ^{1}H polarization of residue i that is transferred via TOCSY and INEPT steps to the amide group of the following $(i+1)$ residue for final $^{1}H^{N}$ detection.

6.2.2 ^{13}C detected experiments

As discussed in paragraph 4.2, the 2D HN and CON spectra provide complementary tools for the investigation of IDPs. Therefore, similarly to the $^{1}H^{N}$-based experiments discussed above, ^{13}C' detection can be extended to higher dimensions, thus providing the required correlation information for sequential resonance assignment of the protein.

A suite of 3D experiments based on ^{13}C' detection is shown in Figure 6.6. The building blocks used for the coherence transfer steps are very similar to those employed in the analogous $^{1}H^{N}$ detected experiments. ^{1}H polarization can be used as a starting source to increase the sensitivity of ^{13}C' detected experiments, while still keeping the experiments "exclusively heteronuclear". In all of the spectra recorded with these experiments, the detection dimension is the ^{13}C' of residue i (C'_i), while the ^{15}N of residue $i+1$ (N_{i+1}) is frequency labeled in one of the indirect dimensions. These C'_i-N_{i+1} correlations are then dispersed in the third dimension by the chemical shift of one (or more) additional nuclear spin(s). In a first set of experiments, CACON and CBCACON [52, 92, 128], the C_i^{α}only or both $C_i^{\alpha}$and C_i^{β} are edited in the third dimension, resulting in one set of correlation peaks per residue, similarly to the sequential $^{1}H^{N}$ detected experiments. In addition, the CCCON-TOCSY experiment [92,128] allows to extend the ^{13}C chemical shift information to all aliphatic side chain carbons of residue i. As a result of this set of experiments, each residue of the protein is associated to a set of C_i^{aliph}-C'_i-N_{i+1} correlations. The number of

aliphatic carbons detected, and their chemical shifts provide information about the amino acid type of the corresponding residue.

The complementary information necessary for sequence-specific assignment is then obtained by acquiring a second set of ^{13}C' detected 3D experiments, named CANCO and CBCANCO [92,128] that correlate the C'$_i$-N$_{i+1}$ to C$^{\alpha}_{i+1}$ and C$^{\alpha}_{i}$ (CANCO) or C$^{\alpha}_{i+1}$ /C$^{\beta}_{i+1}$, and C$^{\alpha}_{i}$/C$^{\beta}_{i}$ (CBCANCO) resulting in two correlation peaks per residue if only C$^{\alpha}$ chemical shifts are detected in the third dimension, or four, if both C$^{\alpha}$ and C$^{\beta}$ are detected. These experiments thus belong to the class of bi-directional correlation experiments. In addition, the 3D (H)N(CA)NCO [92] spectrum can be recorded to obtain correlations involving an additional ^{15}N nuclear spin. This experiment thus belongs to the CO-N-N class of spectra (in analogy to the H-N-N ones introduced above for ^{1}H$^{\mathrm{N}}$ detected experiments). In the 3D (H)N(CA)NCO experiment, the C'$_i$-N$_{i+1}$ pair is further correlated in the third dimension with the ^{15}N of residues i, $i+1$, and $i+2$. Finally, the 3D COCON experiment [129] was developed to correlate the C'$_i$-N$_{i+1}$ pair with carbonyls of residues i, $i+1$, and $i-1$, providing in this way additional complementary information. The MOCCA mixing scheme significantly improves the sensitivity of this experiment [129, 130].

If the whole set of experiments is performed, sequential resonance assignment is obtained by simultaneous matching of C$^{\alpha}$, C$^{\beta}$, C', and N chemical shifts from intra-residue and sequential correlations. The requirement of simultaneous frequency match of four different nuclear spins provides a robust way of solving assignment ambiguities due to resonance overlap in the spectrum of a single nuclear spin species, as often encountered for IDPs.

As also proline residues are detected in these spectra, ^{13}C' direct detection provides an ideal tool for complete sequence-specific assignment of an IDP, provided that sensitivity is sufficient to enable acquisition of the whole set of experiments in a reasonable overall measurement time [52,98]. Our experience on several IDPs characterized by high flexibility indicates that the complete set of 3D experiments described here can be collected with cryogenic probes optimized for ^{13}C detection and yields the complete sequence-specific assignment without any other additional information. For samples of limited solubility ($\sim$100 μM) or using less optimal spectrometer hardware, it is still possible to acquire the most sensitive 2D (CACO, CBCACO, CON) [98] and 3D experiments (CBCACON, CCCON) [52] in order to complement the spectral information obtained from a series of ^{1}H$^{\mathrm{N}}$ detected experiments.

6.2.3 Aliphatic ^{1}H detected experiments

An alternative to the two previously discussed approaches based on the two most well resolved 2D spectra to study IDPs (HN and CON), based either on $^{1}H^{N}$ or $^{13}C'$ direct detection, consists in focusing on $^{1}H^{\alpha}$ detection, as proposed for the assignment of proline-rich regions [131]. More recently, a set of five triple resonance experiments was developed for sequence-specific resonance assignment of IDPs [132,133]. The coherence transfer pathways and correlated frequency information obtained from these experiments are shown in Figure 6.7. The 3D iH(CA)NCO, iHCAN and (HCA)NCO(CA)H experiments were designed for spin system identification, while the 3D H(CA)CON and (HCA)CON(CA)H provide the complementary connectivity information. The $^{1}H^{\alpha}$ signals are clustered in a narrow chemical shift region where also the water proton spins resonate. Therefore resolution in the direct dimension is quite limited, in particular for amino acids of the same kind. Furthermore, excellent water suppression performance is mandatory. Alternatively, the protein can be dissolved in a fully deuterated solvent (99.99% D_2O), which is characterized by higher viscosity with respect to H_2O and causes an increase in the transverse relaxation rates of the nuclear spins in the protein sample. This approach has been successfully applied to an IDP of about 110 amino acids [133]. $^{1}H^{\alpha}$ detected experiments might be considered under conditions where $^{1}H^{N}$ detection of the IDP is not feasible, e.g. at near physiological conditions, and when the protein concentration or the experimental setup does not provide the required sensitivity for ^{13}C detection.

6.3 Experiments for amino acid type editing or selection

In addition to the experiments discussed in the previous sections, which provide sequential correlation information connecting nuclear spins of neighboring residues, information on the amino acid type of a given residue is necessary to complete the sequential resonance assignment of a protein. Amino acid type information may be obtained from NMR data by the selective incorporation of labeled (or unlabeled) amino acids at the protein expression level [134–136], from $^{13}C^{\alpha}$, $^{13}C^{\beta}$ and ^{13}C side-chain chemical shift data [137], or from specifically tailored NMR experiments that exploit the particular spin-coupling topologies and chemical shifts in the different amino acid side chains [138–149]. The latter approach has the advantage that unambiguous amino acid type information is obtained from a series of NMR experiments recorded on a single uniformly $^{13}C/^{15}N$ labeled sample. Such experiments, which often rely of band selective pulses to perturb only specific frequency ranges, benefit from the narrow chemical shift ranges of side chain resonances typical

of IDPs. A selection of the most common experiments available for amino acid type discrimination is presented in this section.

Conceptually we can distinguish between experiments for *amino acid type selection* and those performing *amino acid type editing*. In the first approach (amino acid type selection), only correlation peaks from specific amino acid types are detected, while in the second approach (amino acid type editing), correlation peaks from different classes of amino acids are separated in different NMR subspectra. The two classes of experiments share, as common features, several ingenious ways to distinguish different amino acids exploiting peculiar spectroscopic features of amino acid side chains such as characteristic chemical shifts, coupling topologies, side chain length, etc., as described in the early NMR literature [138, 140–142, 144, 145, 150].

The MUSIC (MUltiplicity Selective In-phase Coherence transfer) strategy proposed by the Oschkinat group [146–148], recently extended to ^{13}C direct detection experiments (CAS-NMR) [151], provides the most complete set of amino acid selective experiments. Different variants of two basic experiments (HNCOCACB and CBCACON) allow to select correlations deriving only from specific amino acids types from the basic 2D HN and CON spectra respectively (Figure 6.8). The resulting spectra are very simple, contributing in this way to reducing the problem of cross-peak overlap. The various pulse schemes mainly differ in the number and lengths of transfer steps required, and thus their overall sensitivity. As a rule of thumb, the experiments for shorter side chains are more sensitive than the corresponding pulse sequences for selecting amino acids with longer side-chains. The different pulse sequence elements designed to select correlations of specific amino acids can also be implemented in "bidirectional" experiments (HNCACB, CBCANCO), providing in this way additional information.

As mentioned before, proline residues are abundant in IDPs, and therefore the identification of proline-neighboring residues is of particular importance for resonance assignment strategies based on $^{1}H^{N}$ detected pulse schemes [38]. Therefore proline-selective experiments based on BEST-TROSY HN(COCAN) and iHN(CAN) pulse sequences, exploiting the particular ^{15}N chemical shift of proline residues in IDPs, have been proposed that yield increased sensitivity with respect to their corresponding MUSIC counterparts, for the identification of proline-neighboring residues [78].

Amino acid type editing differs from amino acid type selection by the fact that correlation peaks for all (observable) residues are detected in the final spectrum, while amino acid type information is obtained by sign encoding. In the simplest type of amino acid type editing, the correlation peaks corresponding to a specific class of amino acids are of

opposite sign with respect to all others, thus allowing simple discrimination of this particular class of amino acids. In addition, if a second (reference) experiment is recorded where all NMR signals have the same sign, peaks belonging to different amino acid type classes can be separated in different spectra at the processing level by addition and subtraction of the two recorded data sets. Such a simple sign encoding has been proposed for many basic correlation experiments involving C^{α} and/or C^{β} carbons, requiring only slight modifications of the original pulse sequences [104, 121, 152]. The same concept of sign encoding can also be used to differentiate between more than 2 classes of amino acids by using HADAMARD spectroscopy [153, 154]. In short, to distinguish between n amino acid classes, HADAMARD NMR spectroscopy requires the recording of n spectra with different sign encoding according to a particular HADAMARD scheme, also called HADAMARD matrix. HADAMARD matrices exist for $n = 2, 4, 8, 12, \ldots$ (multiples of 4). The same HADAMARD matrix employed for encoding is then also used at the processing level to calculate linear combinations of the recorded data set resulting in separate spectra for each amino acid class.This concept (with $n = 8$) is used in the HADAMAC (HADamard-encoded AMino ACid type editing) experiments [155–157]. The sequential HADAMAC (sHADAMAC) experiment, based on the sensitive (H)CBCACONH [107,137] correlation sequence, provides amino acid type information for the residue preceding the detected amide group, and allows to differentiate between seven classes of amino acids: (1) Val-Ile-Ala, (2) Gly, (3) Ser, (4) Thr, (5) Asn-Asp, (6) His-Phe-Trp-Tyr-Cys, (7) Arg-Glu-Lys-Pro-Gln-Met-Leu (Figure 6.9b). The same approach has been recently extended also to intra-residue amino acid type editing (iHADAMAC, Figure 6.9a) [156].

6.4 Automated assignment

The establishment of a vast series of multidimensional NMR experiments has opened the way to the development and improvement of automated assignment programs and to their application to IDPs. The very low chemical shift dispersion and high degeneracy of disordered proteins have always been the major limitations to the performance of such programs. However, in recent years many efforts have been made to increase the robustness and reliability of automatic assignment procedures, and now several algorithms are ready for use. Out of the many, MARS [158], FLYA [159] and TSAR [160] are some examples of the most well-established and promising algorithms that can be exploited to assign IDPs.

In general, the programs require as input only a set of experimental peak lists and the primary sequence of the protein to be assigned. The output can be very different, depending on the particular algorithm employed. For example, the assignment, usually

provided to the user as a text file, can be accompanied by graphical representations and/or additional text files containing, for instance, information about the reliability of the assignment and/or suggestions related to missing or ambiguous assignment. In addition, to facilitate the manual validation of the assignment, some programs provide assignment results also in suitable formats to be directly loaded and read by tools for spectral analysis.

Most importantly, recent improvements have made the input files more and more flexible, accepting peak lists from any combination of multidimensional experiments, provided they are written according to simple but specific set of rules which describe the magnetization transfer pathways employed. Using all the experimental data simultaneously, also incomplete peak lists in which some peaks are missing can be used; furthermore, peak lists containing redundancy of information can be exploited to increase the reliability of the chemical shift assignment.

Finally, automatic resonance assignment algorithms differ for the calculation time necessary to perform the assignment procedure. For example, programs like TSAR are really fast, since they simply compare the submitted chemical shift values and map them along the amino acid sequence of the protein. On the contrary, programs like MARS and FLYA require more computation time since they employ several iterative cycles to find the best correspondence between predicted and submitted chemical shift values, minimizing the propagation of possible initial errors in the assignment and/or in the experimental peak lists. Therefore, the more peak lists are complete and provide complementary information (for example combining $^{1}H^{N}$ and $^{13}C'$ detection), the less is the time required by the assignment algorithms to reach convergence.

7 High-dimensional NMR experiments (nD, with n > 3)

As discussed in the previous paragraphs, various experimental strategies allow improvement of spectral resolution. However, in case of severe spectral overlaps, we recommend to increase the dimensionality of the NMR experiments. Indeed, encoding more than three frequencies into a cross peak observed in multidimensional NMR experiments reduces the chances of accidental signal overlap. In order to take maximum advantage of this approach it is important to maintain highest resolution in all the detected dimensions. To this end, the combination of fast pulsing techniques with non-uniform sampling strategies and the respective processing approaches, described in detail in paragraph 5.1, is crucial to keep experimental time and spectral resolution in a reasonable range.

By high-dimensional NMR we refer to experiments in which three or more indirect dimensions are used to encode chemical shift information within the same experiment. In many cases, to cope with low dispersion of resonance frequencies typical of IDPs, increasing the number of dimensions, provides a unique tool to enhance the resolution in order to accelerate the resonance assignment procedure or to make it possible in difficult cases.

In the next section, high-dimensional (4-5D) NMR experiments are presented, focusing on recent developments tailored for IDPs. A very convenient strategy to simplify the inspection and analysis of the resulting spectra is presented, as well as the most useful high-dimensional experiments tailored for IDPs [161].

7.1 Analysis and inspection of high-dimensional NMR spectra

One of the main barriers to the reception and diffusion of high dimensional NMR experiments has always been the idea that more than three dimensions could be conceptually difficult to handle. Indeed, while it is easy to visualize three-dimensional objects, the same does not hold for objects characterized by a higher number of dimensions. A possible solution consists in breaking them down into combinations of objects of lower dimensionality, as also done when inspecting three-dimensional spectra. Indeed most of the available software applications for the analysis of 3D NMR spectra display only two dimensions at the time, extracted at a specific frequency in the third dimension. In a similar way, higher dimensional NMR spectra can be inspected by visualizing only two dimensions at a time, which are associated to a peak in a 2D or 3D reference spectrum in the case of 4D and 5D experiments, respectively. This procedure is possible as, in most cases, higher dimensional experiments are extensions of 2D or 3D spectra in which the new information is encoded in the additional indirect dimensions. This idea is implemented in the Sparse

Multidimensional Fourier Transform (SMFT) algorithm [162].

SMFT proposes an innovative strategy to examine intuitively four or five-dimensional spectra in a very simple and straightforward way. NUS data are processed with the MFT algorithm (mentioned in paragraph 5.1.2) , but only at some predefined frequencies (hence the prefix "sparse"): in this way, the inspection of the full spectrum is reduced to that of a limited number of spectral regions, which are extremely small with respect to the n-dimensional space of the full spectrum. For convenience, these cross-sections are processed as two-dimensional spectra, being very familiar to almost all NMR users. Therefore, four and five-dimensional spectra appear as a series of 2D spectra to which two or three further frequencies, respectively, are associated. Of course, a prior knowledge of these frequencies is needed: together with the 4D or 5D experiment, respectively a 2D or 3D spectrum, called "basis spectrum", which shares the same correlations with the higher dimensional one, should be acquired to retrieve them. In this way, the dimensions in which the frequencies of the peaks are already known are not processed, whereas the additional two dimensions containing the new information are fully computed, as the positions of the signals are unknown there. This approach is schematically illustrated in Figure 7.1 for 4D and 5D experiments.

SMFT facilitates the analysis of high-dimensionality spectra, since this method provides a way to retrieve all the spectral information content without the need to explore the vast n-dimensional space of the full spectrum. The visualization of the spectrum through inspection of 2D cross-sections provides also further advantages: for example, it enables the use of automatic peak picking tools, which perform in a very reliable and efficient way, as the signals are well-resolved thanks to the high dimensionality of the experiment. In addition, it allows to save great amounts of disk space, since just a small subset of the full spectrum is actually processed: in this way, high digital resolution can be set in all the indirect dimensions.

7.2 Examples of high-dimensional experiments tailored for IDPs

The benefits provided by chemical shift labeling of the heteronuclei including ^{13}C direct detection, ^{1}H longitudinal relaxation enhancement and TROSY, all discussed in the paragraphs 5.2 and 5.3, can be combined with the resolving power of high-dimensional NMR experiments specifically tailored for IDPs. In recent years, a large number of high-dimensional experiments has been developed and profitably used to achieve the sequence-specific assignment of several IDPs of medium and large molecular mass. Their number is expected to increase with future methodological advancements and hardware improve-

ments.

Several of the three-dimensional experiments based on coherence transfer via J-couplings, discussed in paragraph 6.2, can be easily extended to 4D or 5D versions, by explicitly frequency labeling the heteronuclear spins only exploited in 3D experiments for coherence transfer (the nuclear spins generally included in parenthesis in the pulse sequence acronyms). Taking the 3D (H)N(COCA)NH as an example, the coherence is transferred through ^{13}C' and $^{13}C^{\alpha}$ to ^{15}N. With introduction of one or two chemical shift evolution periods, ^{13}C' or/and $^{13}C^{\alpha}$, nuclei can also be frequency labeled, extending the experiment to 4D or 5D, respectively. Another interesting example is provided by the extension to 4D of the 3D HN(CA)CO and HN(CO)CA, commonly used for sequence-specific assignment (4D HNCOCA and 4D HNCACO [163–165]). Based on the 2D HN experiment (H_i^N-N_i), they provide sequential correlations via the ^{13}C' and $^{13}C^{\alpha}$ nuclei for backbone assignment (C'_{i-1}-C^{α}_{i-1} and C^{α}_i-C'_i/C^{α}_{i-1}-C'_{i-1}, respectively) as shown in Figure 7.2a. The former can be used to resolve frequency degeneracy in the 3D HNCO spectrum, whereas the latter relies on $^{13}C^{\alpha}$ and ^{13}C' chemical shifts to establish sequential correlations. An additional 4D iHNCOCA experiment can also be designed [166]. As a general rule, the price to pay for the additional information retrieved in high-dimensional experiments compared to their lower dimensional analogues, consists in a loss in experimental sensitivity of at least a factor of $\sqrt{2}$ for each additional dimension introduced, due to the requirement of phase-sensitive quadrature detection, and because additional spin relaxation occurs during the frequency editing periods.

An overview of the most useful experiments, exemplifying the many possibilities available to retrieve sufficient information for unambiguous sequence-specific assignment is presented hereafter. A summary of the experiments discussed, together with the correlations provided, is reported in Tables A1 and A2 in the Appendix. The high-dimensional experiments can be grouped in different categories according to their lower dimensional basis 2D (HN or CON) and 3D (HNCO or CACON) spectra as well as to the type of information retrieved in the additional indirect dimensions (amino acid type information through ^{13}C side-chain chemical shifts, sequence-specific information through C^{α}, C^{β}, C' and N chemical shifts).

Starting from the most simple and intuitive ones, the 5D HCBCACONH [137, 162, 167], 5D HCCCONH [123,168,169], 4D HCBCACON [170] and 5D HC(CC-TOCSY)CON [123–127, 171] (Figure 7.2b) provide information on the chemical shifts of the ^{1}H and ^{13}C aliphatic spins of each amino acid (i), by exploiting the HNCO (C'_i, N_{i+1}, H^N_{i+1}) or the CON (C'_i, N_{i+1}) as basis spectra, respectively. Therefore, the related 2D cross-

sections that need to be inspected resemble 2D ^{1}H-^{13}C planes that however contain only the correlations observed at the respective frequencies in the basis spectra, providing in this way very clean and straightforward information to identify the residue type and assign side chain ^{1}H and ^{13}C chemical shifts. If necessary, the 5D experiments that exploit the HNCO as basis 3D spectrum can be easily reduced to their 4D analogues (that exploit the 2D HN as a basis spectrum) by simply omitting the evolution of ^{13}C' chemical shifts.

Along the same lines, the 2D HN and CON spectra can be used as basis spectra for high-dimensional experiments that provide in the additional dimensions information to achieve sequence-specific assignment. A variety of different experiments can be included in this class. For example, the 4D HNCACB [172, 173] and 4D HCBCANCO [170, 174] (bidirectional experiments), yield in cross-section the information on ^{1}H and ^{13}C chemical shifts of C^{α} and C^{β} of amino acids (i) and $(i+1)$ for each cross-peak observed in the 2D HN (N_{i+1}, H^{N}_{i+1}) and 2D CON spectra (C'_{i}, N_{i+1}), respectively.

Another class of experiments provides sequential connectivities through ^{15}N chemical shifts. These are particularly useful because, depending on the experimental variant used, they still exploit the same basis spectra (3D HNCO or 2D HN and 2D CON) and provide in the additional dimensions, information on ^{15}N chemical shifts of neighboring residues. This category includes the 4D HN(COCA)NH [120, 175], 4D HN(CA)NNH [173], 5D HN(CA)CONH [162] and 5D HN(COCAN)CONNH [176] ($^{1}H^{N}$ detected) as well as a variety of different experimental variants based on ^{13}C direct detection described below. The $^{1}H^{N}$ detected ones reported above correlate the H^{N}_{i}-N_{i} peak of the 2D ^{1}H-^{15}N HSQC spectrum (the H^{N}_{i}-N_{i}-C'_{i-1} peaks of the 3D HNCO spectrum, in the 5D case) with the $^{1}H^{N}$ and ^{15}N nuclei of residues $i+1$, $i+1/i-1$, $i-1$ and $i+1$, respectively. Instead, when amide protons are merely exploited as a starting source of magnetization, ^{15}N and ^{13}C' resonances can be used to establish sequential correlations, like in the 5D (H)NCO(NCA)CONH [177] (N_{i-1}-C'_{i-2} and N_{i}-C'_{i-1}) and 5D (H)NCO(CAN)CONNH [176] (N_{i}-C'_{i-1} and N_{i+1}-C'_{i}) experiments (Figure 7.2c).

For what concerns the ^{13}C detected analogues, they can be acquired in the 4D mode, using the 2D CON as basis experiment, or extended to the 5D mode by exploiting the C^{α} chemical shift as an additional dimension of the basis experiment. They include 5D NCOCANCO [97], 5D ($H^{N\text{-flip}}$N)CONCACON [178], 5D (HCA)CONCACON [178] and 5D (H)CACON(CA)CON [178] experiments. Sequential correlations are established by exploiting ^{15}N and ^{13}C' frequencies, both retrieved in the 2D cross-section of the experiments ("CON-CON strategy"). The cross-peak observed in cross-section of the experiments reported above correlate N_{i+2}-C'_{i+1}, C'_{i-1}-N_{i}/C'_{i}-N_{i+1}, C'_{i-1}-N_{i}/C'_{i}-N_{i+1} and

C'_{i+1}-N_{i+2}/C'_i-N_{i+1} nuclei, respectively.

For ^{13}C detection, also a series of experiments that rely on CACO as basis 2D spectrum have been proposed. The C^{α}_i-C'_i frequencies of the basis spectrum are correlated to C^{α}_{i+1}-N_{i+1} and C^{α}_{i-1}-N_i or to C^{α}_{i-1}-C'_{i-1} nuclei, respectively through the 4D (H)CANCACO [170] and the 5D CACONCACO [97] experiments. Finally, the 5D $H^{N\text{-flip}}$NCACON [170] and 5D $H^{N\text{-flip}}$NCANCO [170] experiments relate the ^{15}N of residue i with those of residue $i+1$ and $i+2$, respectively (Figure 7.2d).

Recently high-dimensional NMR experiments featuring $^1H^{\alpha}$-start-$^1H^{\alpha}$ detection were also designed. These experients may be useful at temperature and pH conditions in which $^1H^N$ are not detectable and/or in case of short lived or low concentrated samples that partly limit the use of high-dimensional ^{13}C detected experiments. Extensions to higher dimensionality partly counterbalances the low chemical shift dispersion of $^1H^{\alpha}$ in the direct dimension [176].

In conclusion, due to their high resolving power and increased information content, high-dimensional experiments provide a valuable tool to extend the size and complexity of IDPs that can be characterized by NMR at atomic resolution, provided the samples relaxation properties allow for multiple (and often long) coherence transfer times. The combination of the different approaches described in this chapter offers a rich toolbox that can be exploited for the investigation of complex IDPs (Figure 7.3).

8 Conclusions and perspectives

Thanks to recent improvements in NMR instrumental technology, to the array of NMR experiments that have been developed and to their optimization for the specific properties of IDPs a number of tools are available for the atomic resolution structural and dynamic characterization of IDPs.

As of today, IDPs as large as 400 amino acids have been investigated, with two known examples being MAP [179] and Tau [180]; a number of other examples of high resolution investigations of IDPs in the range between 100 and 300 amino acids have appeared in the literature. Many more high resolution NMR experimental investigations of IDPs are expected to be performed in the next future.

Indeed the sequence-specific assignment and initial structural and dynamic characterization through the analysis of chemical shifts and of ^{15}N relaxation rates can readily be achieved through the experiments described in this chapter. Many more experiments can be performed to gain further information as discussed in the next chapters. However, already at this stage the sequence-specific assignment and ^{15}N relaxation measurements are sufficient to describe the overall properties of the IDP in its native state and they can be used as the basis for further investigation of its function by monitoring interactions, post-translational modifications, as well as by taking snapshots of the IDP inside whole cells.

Among the many experimental strategies discussed in detail in this chapter, the optimal one for a specific IDP to be investigated can be readily identified by acquiring a small set of initial 1D and 2D spectra which provide information on the sensitivity and resolution that can be achieved, resulting from the relaxation properties and chemical shift dispersion of the investigated IDP. In particular, the 2D HN and CON spectra are the most suitable to evaluate the most appropriate experimental strategy for sequence-specific assignment. Nowadays, a large panoply of optimized pulse sequences and processing tools are available, as well as user-friendly computational tools for the analysis of the resulting spectra, including higher dimensional ones.

NMR spectroscopy is continuously improving in terms of hardware performance and the availability of new experimental tools. We expect that future progress will aim at further extending the size limit of IDPs that can be investigated at high resolution through NMR, as well as the determination and interpretation of a larger number of observable reporting on structural and dynamic properties of IDPs, and finally the possibility to characterize the behavior of IDPs in different aggregation states, ranging from solution to

solid state to in-cell, through a combination of different NMR techniques.

Finally, speculating on more long term perspectives, development of improved NMR methods to study IDPs is expected to provide a large amount of experimental data on IDPs contributing to our understanding of the molecular basis responsible for their function, filling a gap of about 50 years with respect to our knowledge on the structural and dynamic behavior of folded proteins. This is expected to reveal a much larger number of ways in which proteins communicate in the cell. Other expected outcomes of NMR experimental data on IDPs include the improvement of prediction tools, which still suffer from the bias that they are derived from the missing information in the electron density maps in the X-ray crystallography data.

Appendix

Table A1: High-multidimensional ^{1}H detected experiments for backbone and side-chains resonance assignment

^{1}**Hdetected experiments**			
Experiments for spin-system identification			
Dimensionality	**Experiment**	**Correlations observed**	**References**
3	HN(CO)CA	H_i^N-N_i-C_{i-1}^{α}	[78, 104, 106]
3	HN(CO)CACB	H_1^N-N_i-$C_{i-1}^{\alpha/\beta}$	[78, 107, 108]
3	(H)C(CC-TOCSY) (CO)NH	C_{i-1}^{ali}-N_i-H_i^N	[123–125]
3	iHNCA	H_i^N-N_i-C_i^{α}	[78, 114, 115]
3	iHNCACB	H_i^N-N_i-$C_i^{\alpha/\beta}$	[78, 114, 115]
3	iHCAN	H_i^{α}-C_i^{α}-N_i	[132]
4	(H)CBCACONH	$C_{i-1}^{\alpha/\beta}$-C'$_{i-1}$-N_i-H_i^N	[107, 137]
5	HCBCACONH	$H_i^{\alpha/\beta}$-$C_i^{\alpha/\beta}$-C'$_{i-1}$-N_i-H_i^N	[137, 162, 167]
5	HNCOCACB	H_i^N-N_i-C'$_{i-1}$-C_{i-1}^{α}-C_{i-1}^{β}	[177, 181]
5	HC(CC-TOCSY)CONH	H_{i-1}^{ali}-C_{i-1}^{ali}-C'$_{i-1}$-N_i-H_i^N	[123–127, 171]
Experiments for sequential assignment			
Dimensionality	**Experiment**	**Correlations observed**	**References**
3	HNCO	H_i^N-N_i-C'$_{i-1}$	[78, 103–105]
3	HNCA	H_i^N-N_i-C_{i-1}^{α}& H_i^N-N_i-C_i^{α}	[46, 103, 104]
3	HNCACB	H_i^N-N_i-$C_{i-1}^{\alpha/\beta}$& H_i^N-N_i-$C_i^{\alpha/\beta}$	[46, 109, 110]
3	HN(CA)CO	H_i^N-N_i-C'$_{i-1}$& H_i^N-N_i-C'$_i$	[46, 111–113]
3	(H)N(COCA)NH	N_{i+1}-N_i-H_i^N	[118, 120, 121]
3	(H)N(CA)NNH	N_{i-1}-N_i-H_i^N& N_{i+1}-N_i-H_i^N	[118, 119]

Dimensionality	Experiment	Correlations observed	References
3	iH(CA)NCO	H_i^{α}-N_i-C'$_i$	[132]
3	(HCA)NCO(CA)H	H_i^{α}-C'$_i$-N_{i+1}	[133]
3	H(CA)CON	H_i^{α}-C'$_i$-N_{i+1}	[132]
3	(HCA)CON(CA)H	H_i^{α}-N_i-C'$_i$ & H_i^{α}-N_{i+1}-C'$_i$	[133]
4	HCACON	H_i^{α}-C_i^{α}-C'$_i$-N_{i+1}	[182]
4	HNCOCA	H_i^{N}-N_i-C'$_{i-1}$-C_{i-1}^{α}	[163–165]
4	HNCACO	H_i^{N}-N_i-C_i^{α}-C'$_i$& H_i^{N}-N_i-C_{i-1}^{α}-C'$_{i-1}$	[164, 165]
4	HNCO,CA	H_i^{N}-N_i-C'$_{i-1}$-C_{i-1}^{α}& H_i^{N}-N_i-C'$_{i-1}$-C_i^{α}	[166]
4	HNCACB	H_i^{N}-N_i-C_{i-1}^{α}-C_{i-1}^{β}& H_i^{N}-N_i-C_i^{α}-C_i^{β}	[172]
4	HCBCANH	$H_i^{\alpha/\beta}$-$C_i^{\alpha/\beta}$-N_i-H_i^{N}& $H_{i-1}^{\alpha/\beta}$-$C_{i-1}^{\alpha/\beta}$-N_i-H_i^{N}	[173]
4	HACANH	H_i^{α}-C_i^{α}-N_i-H_i^{N}& H_{i-1}^{α}-C_{i-1}^{α}-N_i-H_i^{N}	[183, 184]
4	HNCAHA	H_i^{N}-N_i-C_i^{α}-H_i^{α}& H_{i+1}^{N}-N_{i+1}-C_i^{α}-H_i^{α}	[165]
4	HACA(CO)NH	H_{i-1}^{α}-C_{i-1}^{α}-N_i-H_i^{N}	[183]
4	HN(CO)CAHA	H_{i+1}^{N}-N_{i+1}-C_i^{α}-H_i^{α}	[165]
4	(H)CACO(CA)NH	C'$_i$-C_i^{α}-N_i-H_i^{N}& C'$_{i-1}$-C_{i-1}^{α}-N_i-H_i^{N}	[185]
4	HN(COCA)NH	H_{i+1}^{N}-N_{i+1}-N_i-H_i^{N}& H_i^{N}-N_i-N_i-H_i^{N}	[120, 175]
4	HN(CA)NH	H_{i+1}^{N}-N_{i+1}-N_i-H_i^{N}& H_i^{N}-N_{i-1}-N_{i-1}-H_i^{N}	[173]
4	HN(CO)CA(CON)CA	H_i^{N}-N_i-C_{i-1}^{α}-C_{i-1}^{α}& H_i^{N}-N_i-C_{i-1}^{α}-C_i^{α}	[186]
4	HNCO(N)CA	H_i^{N}-N_i-C'$_{i-1}$-C_{i-1}^{α}& H_i^{N}-N_i-C'$_{i-1}$-C_i^{α}	[186]

Dimensionality	Experiment	Correlations observed	References
5	HACACONH	H^{α}_{i-1}-C^{α}_{i-1}-C'_{i-1}-N_i-H^{N}_i	[187–189]
5	HACA,CONH	H^{α}_{i}-C^{α}_{i}-C'_{i}-N_i-H^{N}_i& H^{α}_{i-1}-C^{α}_{i-1}-C'_{i-1}-N_i-H^{N}_i	[190]
5	HACA(N)CONH	H^{α}_{i}-C^{α}_{i}-C'_{i-1}-N_i-H^{N}_i& H^{α}_{i-1}-C^{α}_{i-1}-C'_{i-1}-N_i-H^{N}_i	[177]
5	(HACA)CON(CA)CONH	C'_{i-2}-N_{i-1}- C'_{i-1}-N_i-H^{N}_i& C'_{i-1}-N_i-C'_{i-1}-N_i-H^{N}_i	[177]
5	HCBCA(CAN)CONH	$H^{\alpha/\beta}_{i-1}$-$C^{\alpha/\beta}_{i-1}$- C'_{i-1}-N_i-H^{N}_i& $H^{\alpha/\beta}_{i}$-$C^{\alpha/\beta}_{i}$-C'_{i-1}-N_i-H^{N}_i	[167]
5	HN(CA)CONH	H^{N}_{i-1}-N_{i-1}- C'_{i-1}-N_i-H^{N}_i& H^{N}_{i}-N_i-C'_{i-1}-N_i-H^{N}_i	[162]
5	(H)NCO(NCA)CONH	N_{i-1}-C'_{i-2}- C'_{i-1}-N_i-H^{N}_i& N_{i-1}-C'_{i-1}-C'_{i-1}- N_i-H^{N}_i	[177]
5	(H)NCO(CAN)CONNH	N_{i+1}-C'_{i}-H^{N}_i	[176]
5	HN(COCAN)CONNH	H^{N}_{i+1}-N_{i+1}-C'_{i-1}-N_i-H^{N}_i	[176]
5	(HACA)CON(CACO) NCO(CA)HA	C'_{i-1}-N_i-N_{i+1}- -C'_{i}-C^{α}_{i}	[176]

Table A2: High-multidimensional ^{13}C detected experiments for backbone and side-chains resonance assignment

^{13}C detected experiments			
Experiments for spin-system identification			
Dimensionality	**Experiment**	**Correlations observed**	**References**
3	(H)CACON	C_i^{α}-N_{i+1}-C'$_i$	[92, 128]
3	(H)CBCACON	$C_i^{\alpha/\beta}$-N_{i+1}-C'$_i$	[92, 128]
3	(H)C(CC-TOCSY)CON	C_i^{ali}-N_{i+1}-C'$_i$	[92, 128]
4	HCBCACON	$H_i^{\alpha/\beta}$-$C_i^{\alpha/\beta}$-N_{i+1}-C'$_i$	[170, 174]
4	HC(CC-TOCSY)CON	H_i^{ali}-C_i^{ali}-N_{i+1}-C'$_i$	[170]
5	HC(CC-TOCSY)CACON	H_i^{ali}-C_i^{ali}-C_i^{α}-N_{i+1}-C'$_i$	[179]
Experiments for sequential assignment			
Dimensionality	**Experiment**	**Correlations observed**	**References**
3	(H)CANCO	C_i^{α}-N_{i+1}-C'$_i$& C_{i+1}^{α}-N_{i+1}-C'$_i$	[92, 128]
3	(H)CBCANCO	$C_i^{\alpha/beta}$-N_{i+1}-C'$_i$& $C_{i+1}^{\alpha/beta}$-N_{i+1}-C'$_i$	[92, 128]
3	($H^{N\text{-flip}}$)N(CA)NCO	N_i-N_{i+1}-C'$_i$& N_{i+2}-N_{i+1}-C'$_i$	[92, 170]
3	(HCA)NCACO	N_i-C_i^{α}-C'$_i$& N_{i+1}-C_i^{α}-C'$_i$	[170]
3	COCON	C'$_{i-1}$-N_{i+1}-C'$_{i-}$& C'$_{i+1}$-N_{i+1}-C'$_{i-}$	[129]
4	HCBCANCO	$H_i^{\alpha/\beta}$-$C_i^{\alpha/\beta}$-N_{i+1}-C'$_i$& $H_{i+1}^{\alpha/\beta}$-$C_{i+1}^{\alpha/\beta}$-N_{i+1}-C'$_i$	[170, 174]
4	($H^{N\text{-flip}}$)NCANCO	N_i-C_i^{α}-N_{i+1}-C'$_i$& N_{i+2}-C_{i+1}^{α}-N_{i+1}-C'$_i$	[170]
4	($H^{N\text{-flip}}$)NCACON	N_i-C_i^{α}-N_{i+1}-C'$_i$& N_{i+1}-C_i^{α}-N_{i+1}-C'$_i$	[170]

Dimensionality	Experiment	Correlations observed	References
4	(H)CANCACO	C^{α}_{i-1}-N_i-C^{α}_i-C'_i& C^{α}_{i+1}-N_{i+1}-C^{α}_i-C'_i	[170]
5	$H^{N\text{-flip}}$NCANCO	H^{N}_i-N_i-C^{α}_i-N_{i+1}-C'_i& H^{N}_{i+2}-N_{i+2}-C^{α}_{i+1}-N_{i+1}-C'_i	[170]
5	$H^{N\text{-flip}}$NCACON	H^{N}_i-N_i-C^{α}_i-N_{i+1}-C'_i& H^{N}_{i+1}-N_{i+1}-C^{α}_{i+1}-N_{i+1}-C'_i	[170]
5	(H)NCOCANCO	N_{i+2}-C'_{i+1}-C^{α}_{i+1}-N_{i+1}-C'_i	[97]
5	(H)CACONCACO	C^{α}_{i-1}-C'_{i-1}-N_i-C^{α}_i-C'_i	[97]
5	($H^{N\text{-flip}}$N)CONCACON	C'_{i-1}-N_i-C^{α}_i-N_{i+1}-C'_i& C'_i-N_{i+1}-C^{α}_i-N_{i+1}-C'_i	[178]
5	(HCA)CONCACON	C'_{i-1}-N_i-C^{α}_i-N_{i+1}-C'_i& C'_i-N_{i+1}-C^{α}_i-N_{i+1}-C'_i	[178]
5	(H)CACON(CA)CON	C^{α}_i-C'_i-N_{i+1}-C'_{i+1}-N_{i+2}-& C^{α}_i-C'_i-N_{i+1}-C'_i-N_{i+1}	[178]

Tables

Table 1.1: Gyromagnetic ratios and natural abundance of nuclei important for NMR studies of proteins.

Isotope	Spin I	gyromagnetic ratio $\gamma_n(10^6 \text{rads}^{-1}\text{T}^{-1})$	gyromagnetic ratio/2π $\gamma_n/2\pi(\text{MHzT}^{-1})$	natural abundance isotope %
^{1}H	1/2	267,512	42,576	99,98
^{13}C	1/2	67,262	10,705	1,108
^{15}N	1/2	-27,116	-4,316	0,37

Table 5.1: Ranges of predicted amide proton solvent exchange rates (k_{ex}) (predictions were performed with SPHERE program [43,44]) and their average over all residues k_{ex} are listed as well as the ranges and averages of the apparent T_1 relaxation time constants measured at 800 MHz as shown in Figure 5.9.

	NS5A	BASP-1	α-synuclein	PV Core
Conditions				
pH	6.5	2.0	7.4	7.5
T(C)	5	5	15	5
Predictions				
range of k_{ex}	$7.6 \times 10^{-2}-$	$1.9 \times 10^{-4}-$	$1.8 \times 10^{-1}-$	$8.7 \times 10^{-2}-$
(s^{-1})	$1.2 \times 10^{+2}$	1.3×10^{-3}	$2.1 \times 10^{+2}$	$1.6 \times 10^{+2}$
$\bar{k}_{ex}(s^{-1})$	1.5	3.7×10^{-4}	$1.7 \times 10^{+1}$	$1.1 \times 10^{+1}$
NMR data				
$\bar{T}_{1,hard}$(s)	0.92	0.91	2.27	1.46
$\bar{T}_{1,WFB}$(s)	0.70	0.89	0.09	0.14
$\bar{T}_{1,sel}$(s)	0.21	0.31	0.06	0.09

Figures

Figure 2.1. Schematic illustrations of differences between structured (**a**) and intrinsically disordered (**b**) proteins (for clarity reasons the molecules of water were omitted in the ensemble of IDP conformers), and energy landscapes of well-folded proteins in native state (**c**) and IDP (**d**).

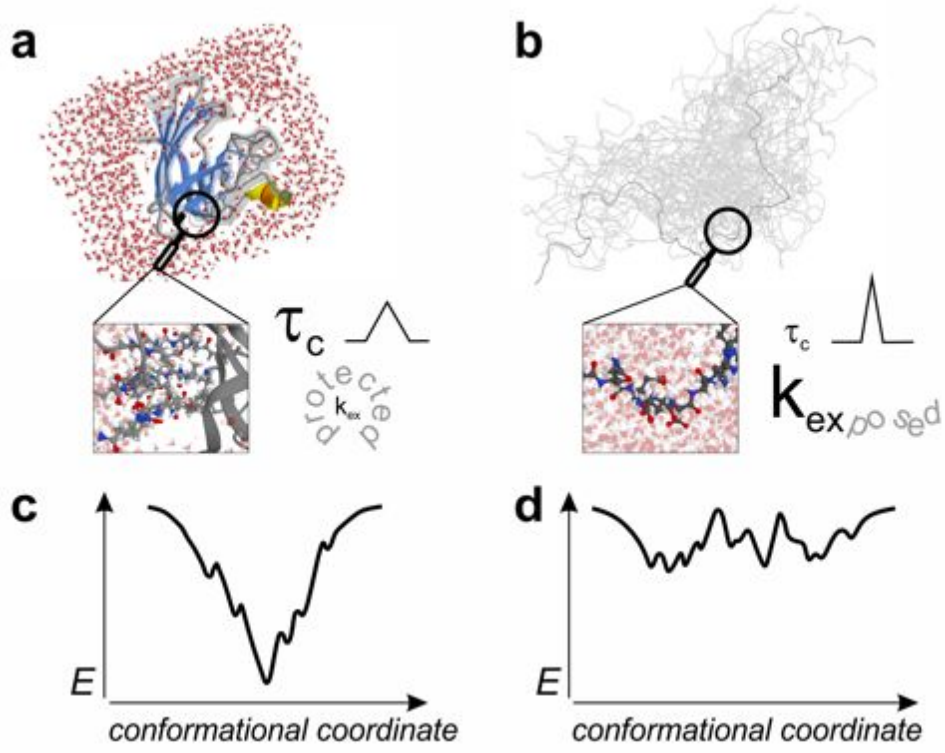


Figure 2.2. The different chemical shift dispersion of the structured and intrinsically disordered proteins is demonstrated by 2D ^{1}H-^{15}N correlation spectra acquired on two proteins of the similar size, but characterized by the different structural properties. **a** The HSQC spectrum of structured Cu(I)Zn(II) superoxide dismutase (1.5 mM sample in 20 mM phosphate buffer, pH 5.0, at 298 K; PDB code: 1MFM). **b** The HSQC spectrum of intrinsically disordered α-synuclein proteins (1.0 mM sample in 20 mM phosphate, pH 6.4, 0.5 mM EDTA, 200 mM NaCl, at 285.5 K). The experiments were acquired at 700 MHz Bruker AVANCE spectrometer equipped with CPTXI probe.

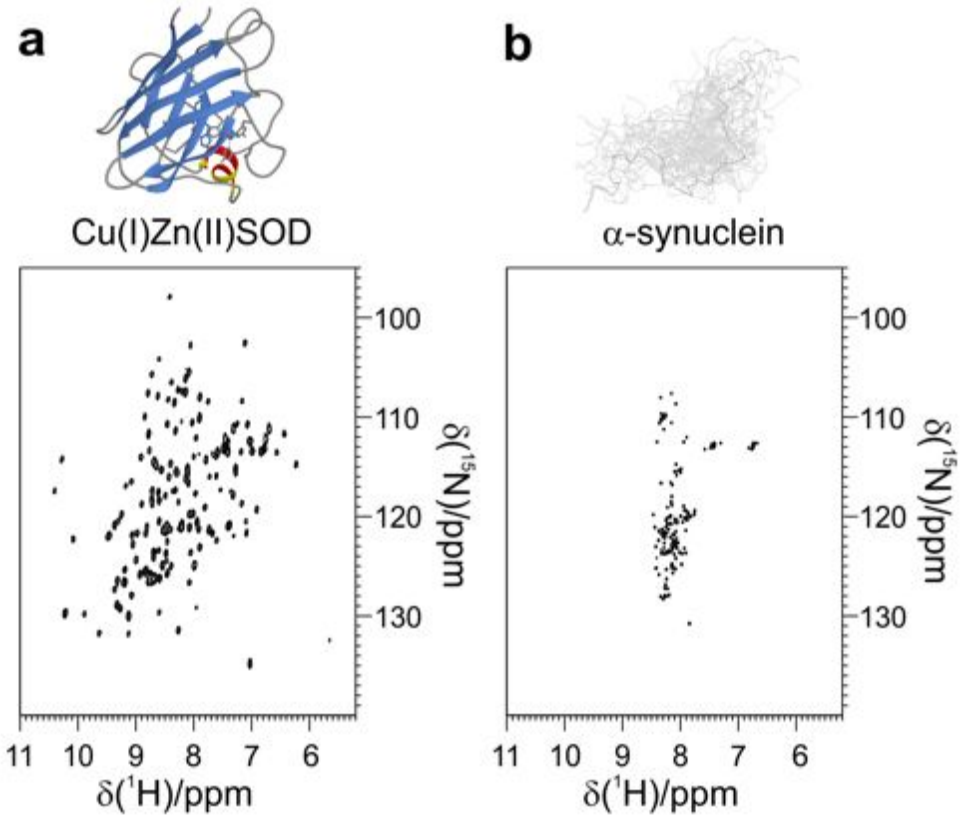


Figure 2.3. **a** Histograms of the C^{α} and C^{β} secondary chemical shift distribution in α-helix and β-sheet.(adopted from [7]) **b** Identification of the secondary structure elements based on the secondary chemical shifts is shown on the examples of structured metallochapenone Atx1 [191], partially structured oxidized form of the copper chaperone Cox 17 [192] and its unfolded reduced form [193].

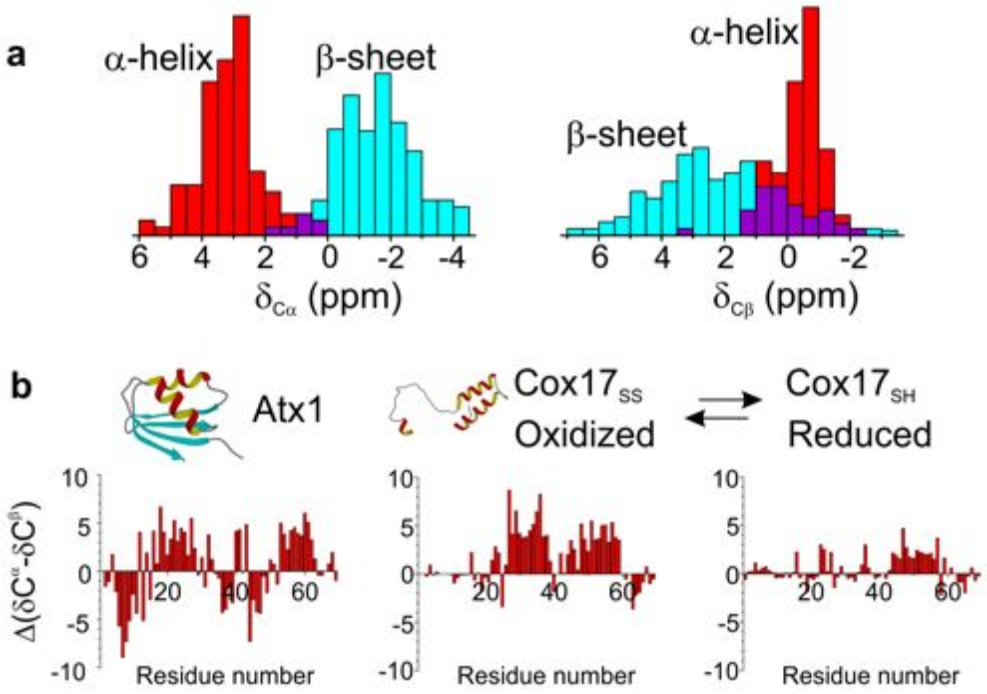


Figure 2.4. **a** Variations of the steady-state NOE with the correlation time for the ^{1}H-^{15}N pair in a field of 11.74 T. **b** The graphs shows ^{1}H-^{15}N heteronuclear NOE for the majority of the residues of the structured Atx1 (1 mM sample, 100 mM phosphate buffer, pH 7.0, at 298 K) [191], partially structured and unfolded form of Cox 17 proteins (2 mM sample, 50 mM phosphate, pH 7, at 298 K; sample with unfolded form was reduced by 20 mM DTT) [192, 193]. The experiments were acquired at 500 MHz Bruker AVANCE spectrometer equipped with CPTXI probe.

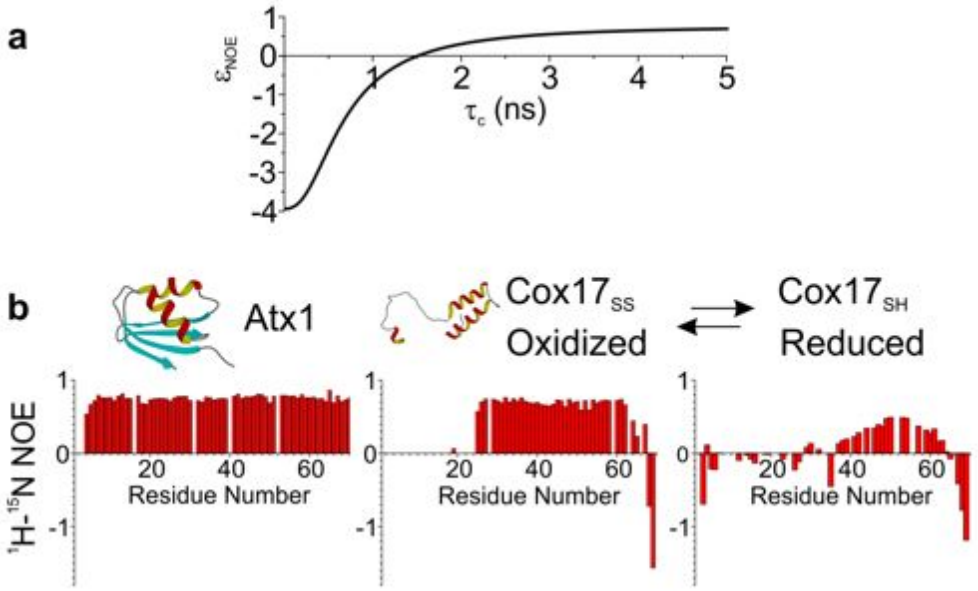


Figure 3.1. The difference of the ^{1}H, ^{13}C and ^{15}N chemical shift dispersion of the IDPs is illustrated on the 1D spectra of α-synuclein. The experiments were acquired at 700 MHz Bruker AVANCE spectrometer equipped with CPTXO probe on 1 mM α-synuclein sample (20 mM phosphate buffer, pH 6.4, 200 mM NaCl, 0.5 mM EDTA, at 285.5 K).

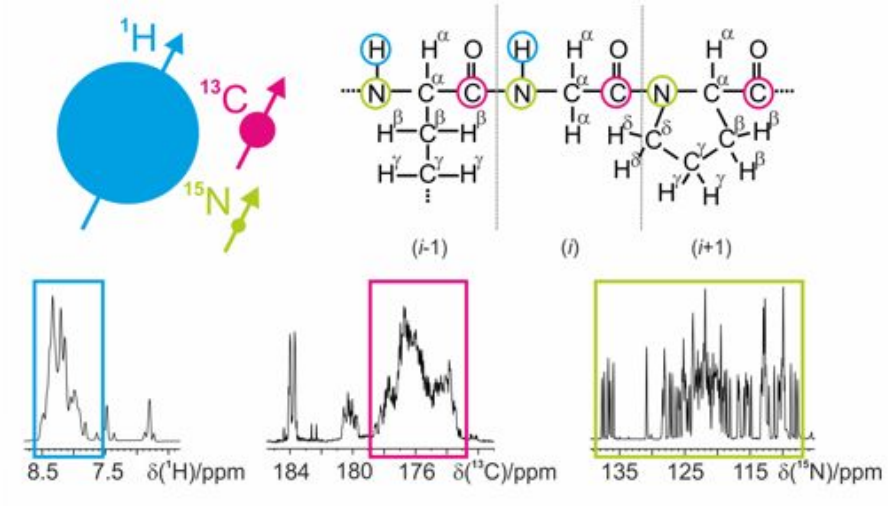


Figure 3.2. ^{15}N amide T_2 relaxation time constants of NS5A measured as function of the amino acid sequence at 278 K, at 18.8 T field. Red cylinders indicate regions adopting transient α-helical secondary structure.

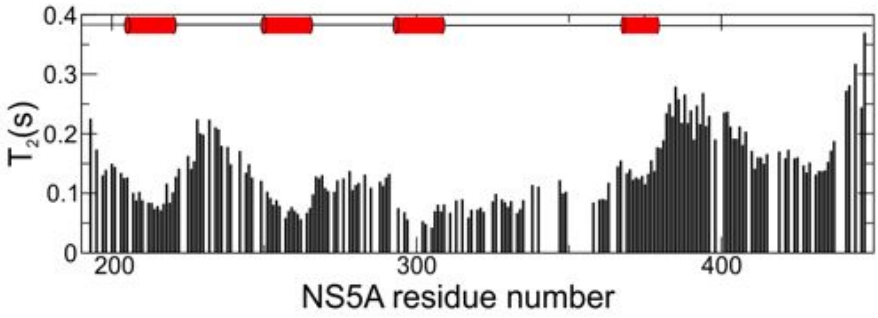


Figure 4.1. The ^{1}H-^{13}C HSQC spectra acquired at 950 MHz Bruker AVANCE spectrometer equipped with CPTXI probe on 1 mM α-synuclein sample (20 mM phosphate buffer, pH 6.4, 200 mM NaCl, 0.5 mM EDTA, at 285.5 K).

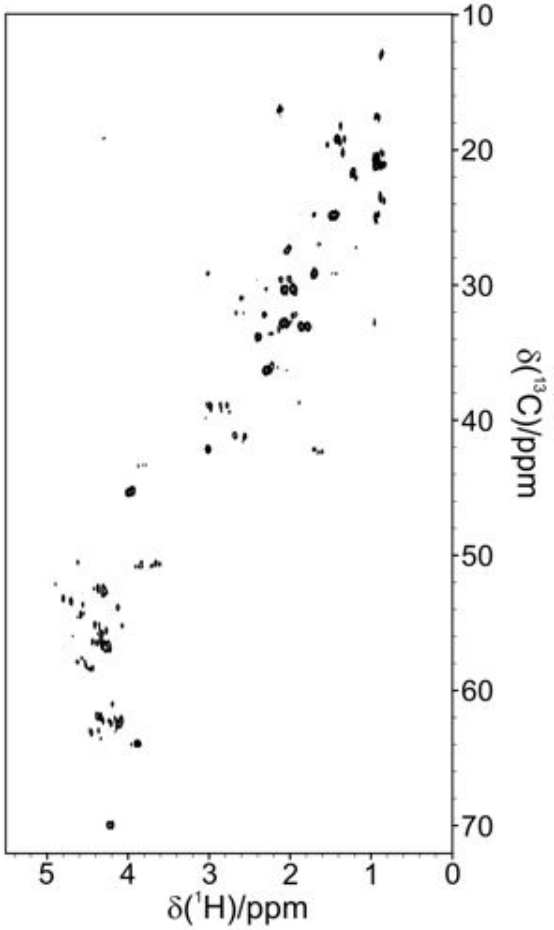


Figure 4.2. The schematic illustration (top) of the correlations observed in the basic 2D ^{1}H-^{15}N HSQC and ^{13}C-^{15}N CON-IPAP experiments. The ^{1}H-^{15}N HSQC (left) and ^{13}C-^{15}N CON-IPAP (right) spectra acquired on the ^{13}C, ^{15}N-labeled sample of the intrinsically disordered human securing (0.7 mM sample in 25 mM phosphate buffer, pH 7.2, 150 mM KCl, 10 mM 2-mercaptoethanol, at 283 K) [194]. The experiments were acquired at Bruker AVANCE 700 MHz spectrometer equipped with the CPTXO probe.

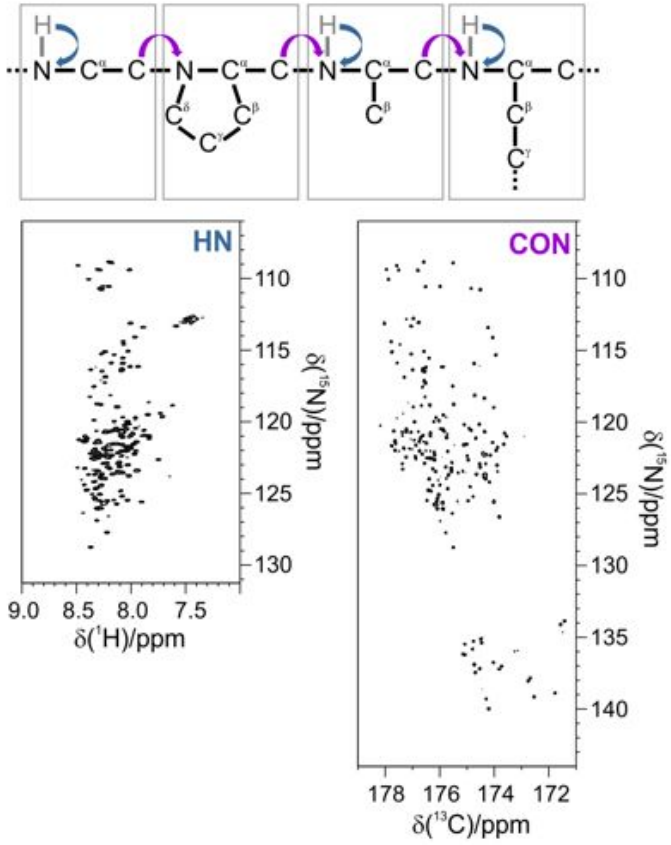


Figure 4.3. The 2D ^{1}H-^{15}N HSQC (**a**) and ^{13}C-^{15}N CON-IPAP (**b**) spectra acquired on 1 mM α-synuclein sample (20 mM phosphate buffer, pH 7.4, 200 mM NaCl, 0.5 mM EDTA) at different temperatures, from left to right: 285.7 K, 295.5 K and 304.8 K. Each experiment was acquired with one scan per increment and with the same resolution (in Hz).

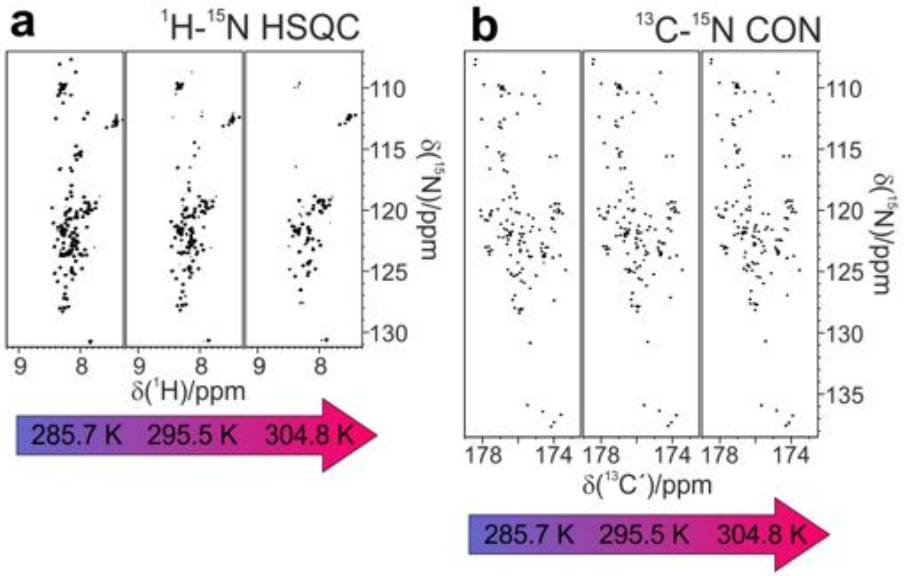


Figure 4.4. The 2D ^{1}H-^{15}N SOFAST-HMQC (**a**, **b**) and ^{13}C-^{15}N H$^{\alpha-\text{flip}}$CON (**c**, **d**) spectra acquired on α-synuclein overexpressed in *E. coli* cells (**b**, **d**) and on cell lysates (**a**, **c**). The experiments were acquired at Bruker AVANCE 700 MHz spectrometer equipped with the CPTXO probe, the acquisition time of SOFAST-HMQC was 13 minutes and of H$^{\alpha-\text{flip}}$CON 44 minutes.

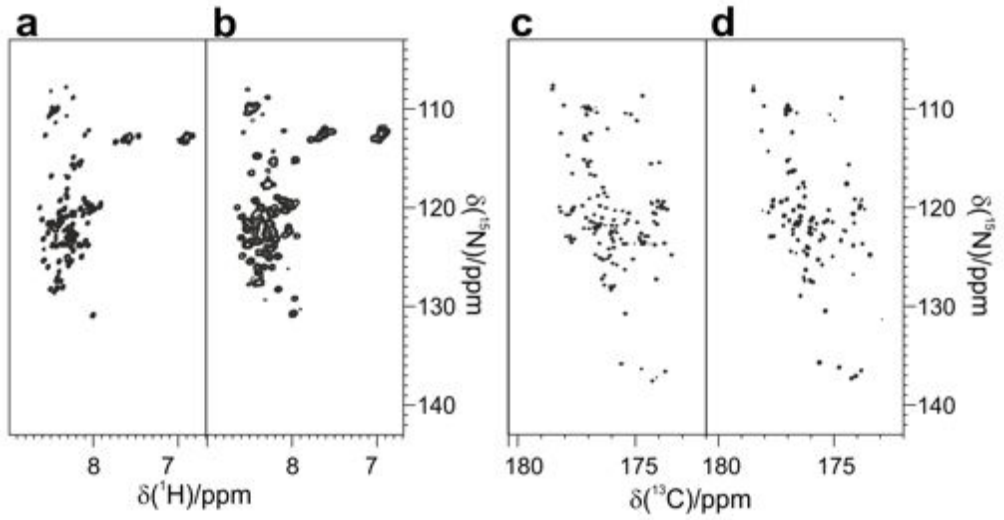


Figure 5.1. Schematic representation of the multi-dimensional NMR spectra (left), and the increasing number of experimental repetitions and thus longer acquisition times (right).

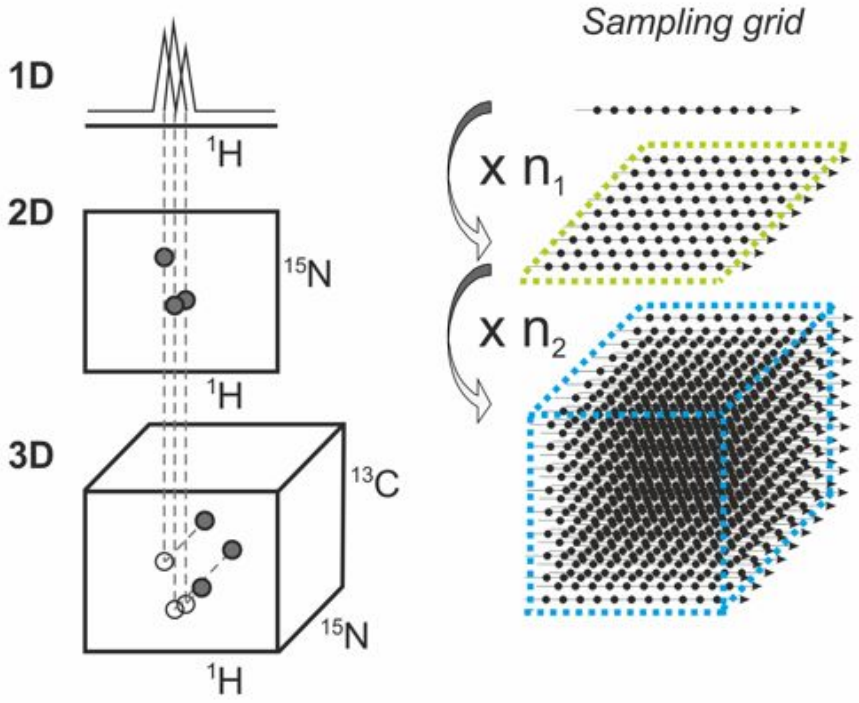


Figure 5.2. Common pulse sequence implementations of I-spin editing during an incremented time variable t_1: **a**, conventional real-time; **b**, optimized real-time; **c** and **d** constant-time (CT); **e** semi-CT editing.

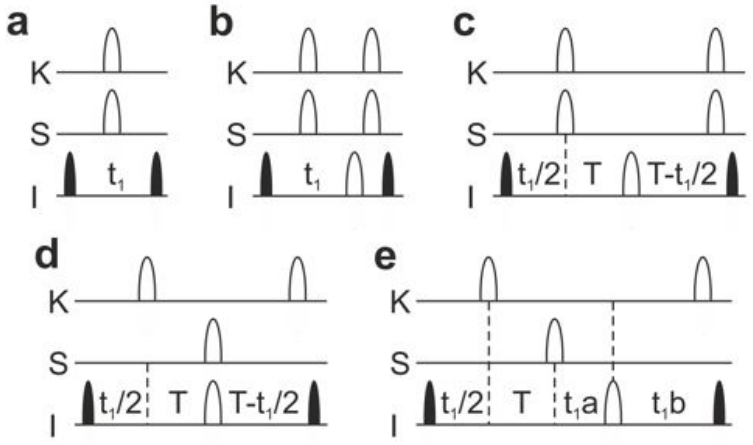


Figure 5.3. Expansion of ^{1}H-^{13}C$^{\alpha}$ plane from a BEST-TROSY HN(CO)CA spectrum of NS5A recorded with real-time (**a**) and CT evolution (**b**). In both cases the acquisition time in the indirect ^{13}C$^{\alpha}$ dimension was 20 ms with 200 points in a spectral width of 30 ppm.

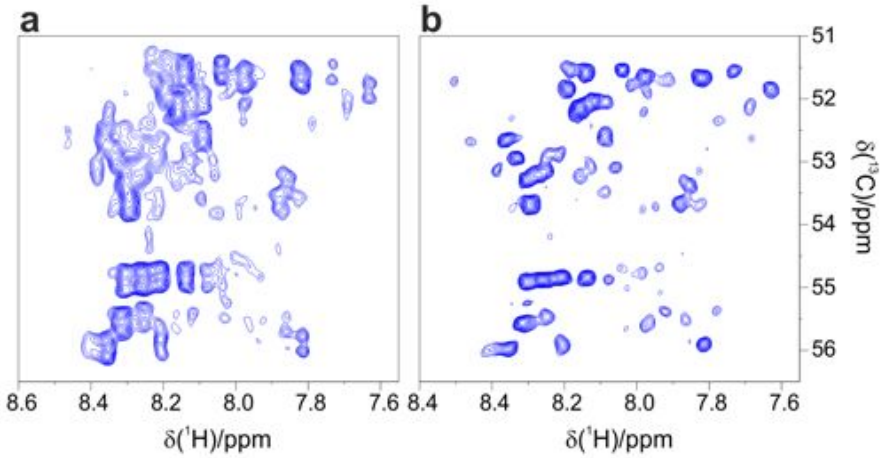


Figure 5.4. Expansion of ^{1}H-^{15}N plane from BEST-TROSY HNCA of NS5A acquired with (**a**) CT editing using the maximum number of increments (AT = 24.5 ms) and (**b**) semi-CT evolution (AT = 53 ms). In both cases the spectral width in the ^{15}N dimension was 30 ppm.

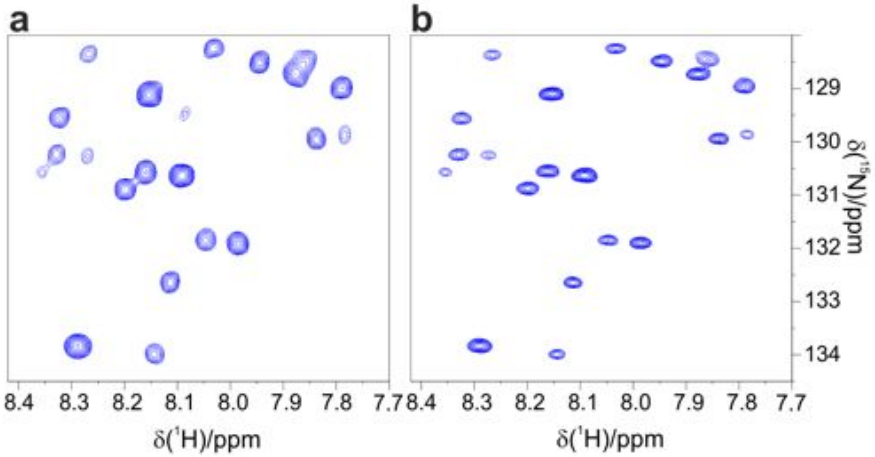


Figure 5.5. Examples of different sampling schemes to speed up the acquisition of a 3D experiment. **a** Schematic illustration of a typical three-dimensional H-X-Y correlation experiment and the conventional time-domain sampling grid required to build the two indirect dimensions. Each point on the grid corresponds to a single repetition of the basic pulse sequence. **b** Some NUS schemes are shown. All the sampling patterns can be employed to reduce the experimental time, since a smaller number of points with respect to those of the conventional sampling grid are measured. The linear under sampling pattern can be used to fold chemical shifts in an advantageous manner, whereas radial sampling is generally employed in Automated Projection Spectroscopy (APSY). Instead, random non-uniform sampling schemes are most commonly used when non-linear methods for spectral reconstruction, like MDD or MFT, are used.

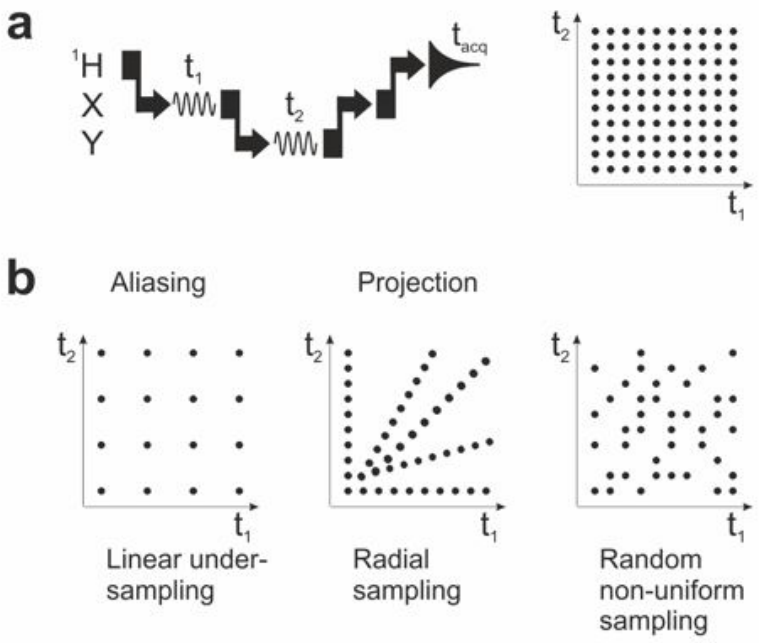


Figure 5.6. Small spectral region of (**a**) ^{1}H-^{15}N coupled HSQC and (**b**) TROSY spectrum of NS5A recorded at 950 MHz spectrometer. The spectral region displays correlation peaks for two NS5A residues. The TROSY effect is less pronounced for the left residue, located in highly flexible region of the IDP, compared to the right one that is part of a peptide segment with a high propensity to form -helical structure. In both cases the acquisition time in indirect dimension was 44 ms with 256 points in a spectral width of 30 ppm.

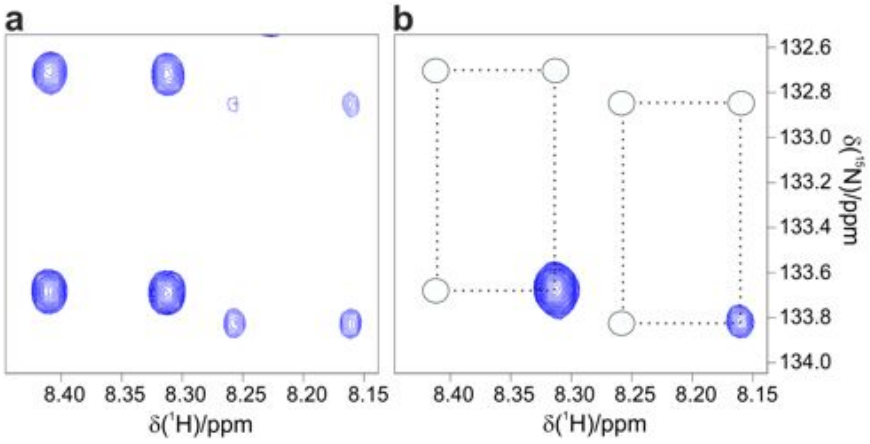


Figure 5.7. A schematic drawing of a NMR experiment.

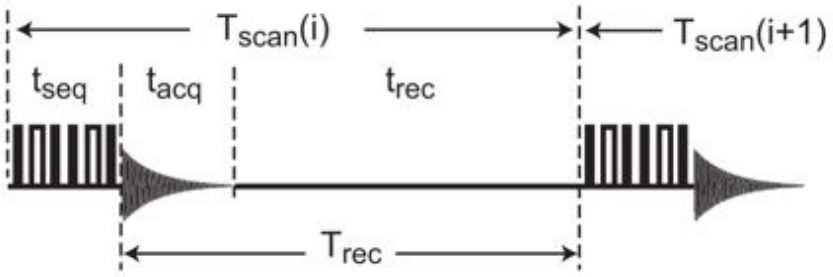


Figure 5.8. Apparent ^{1}H T_1 relaxation time constants measured by inversion recovery experiments. Amide proton-selective (red), water-flip-back (green) and non-selective (black) inversions were used. The four samples and the conditions were: **a**, NS5A D2D3 protein (pH 6.5, 278 K); **b**, BASP1 (pH 6.5, 278 K); **c**, α-synuclein (pH 7.4, 288 K); **d**, PV core (pH 7.5, 278 K).

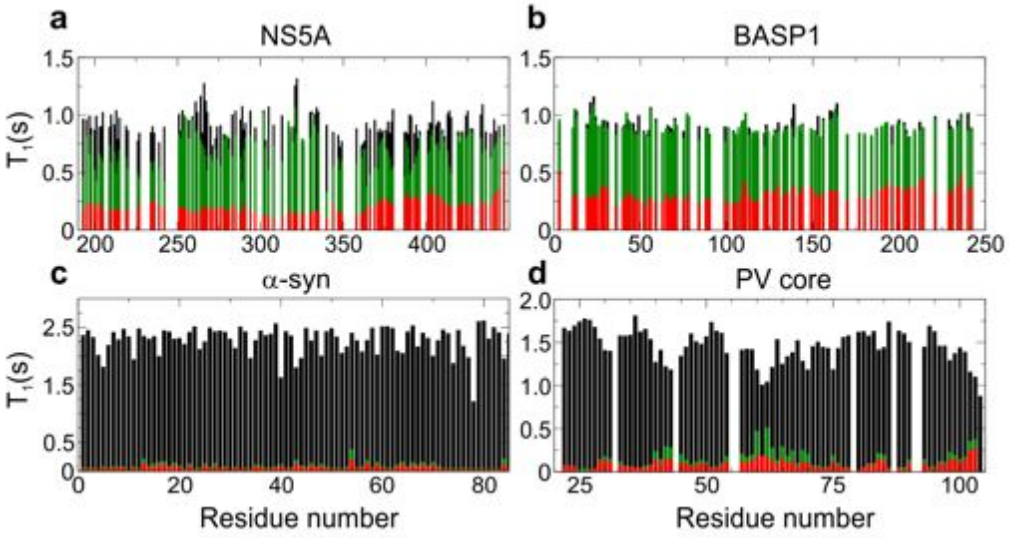


Figure 5.9. 2D ^{1}H-^{15}N BEST-TROSY spectra recorded on two large IDPs (∼270 residues), (**a**) NS5A D2D3 (pH 6.5) and (**b**) BASP1 (pH 2.0) at 278 K

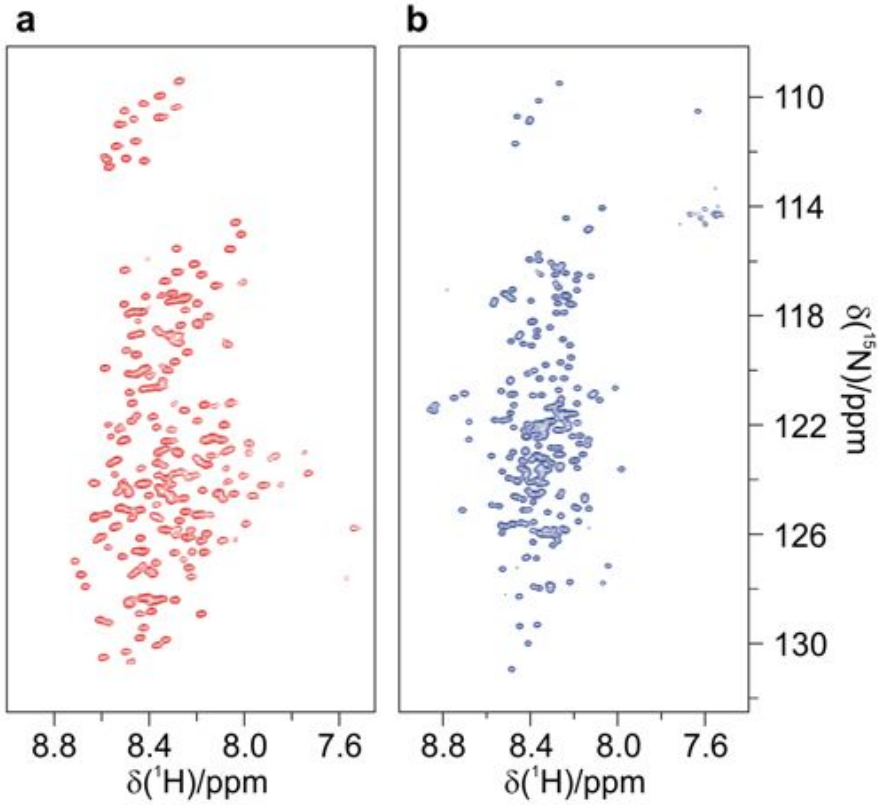


Figure 5.10. Example of 2D ^{13}C detected CON spectra without and with homonuclear decoupling. The experiments were acquired on α-synuclein (1 mM sample in 20 mM phosphate buffer, pH 6.4, 200 mM NaCl, 0.5 mM EDTA, at 285.5 K) on 700 MHz Bruker AVANCE equipped with CPTXO probe.

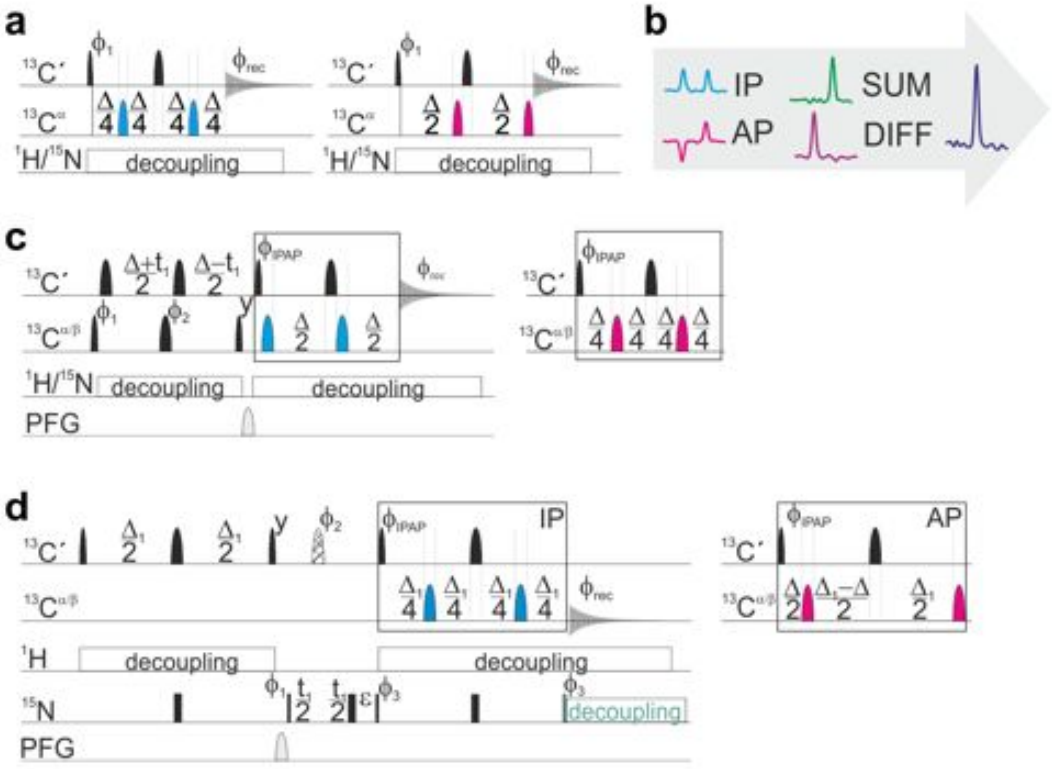


Figure 5.11. ^{13}C detected 1D and 2D experiments with implemented IPAP decoupling sequence. Implementation of $^{13}C'$-$^{13}C^{\alpha}$ IPAP virtual decoupling building blocks in the 1D mode (**a**) and 2D mode (**c**, **d**). Band-selective ^{13}C pulses are denoted by rounded rectangles (narrow and wide ones represent 90° and 180° pulses, respectively). The pulses are applied along the x-axis unless noted differently. **b** The schematic illustration of the post-acquisition processing for obtaining decoupled spectra.

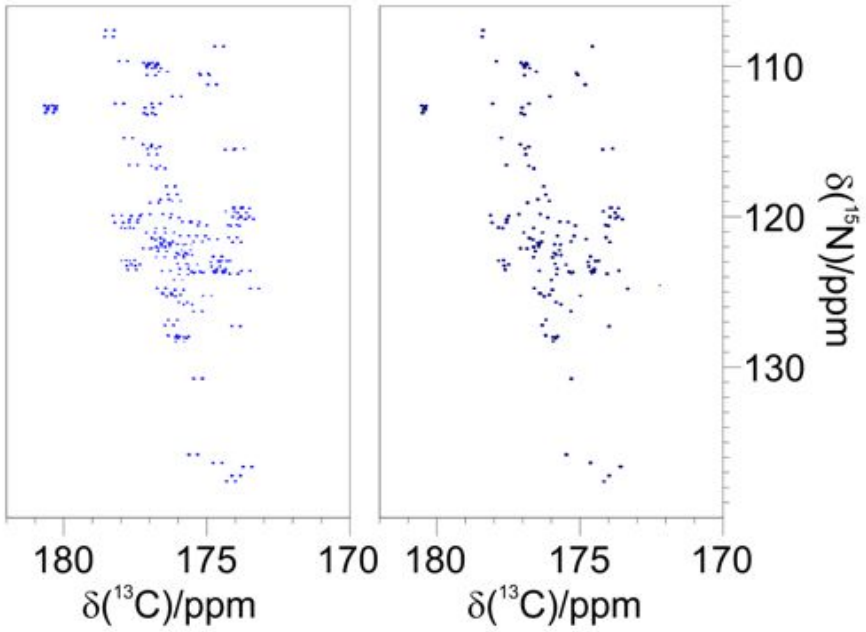


Figure 5.12. The 2D ^{13}C-^{15}N H$^{\text{N-BEST}}$CON spectrum of α-synuclein overexpressed in *E. coli* cells acquired in 20 minutes.

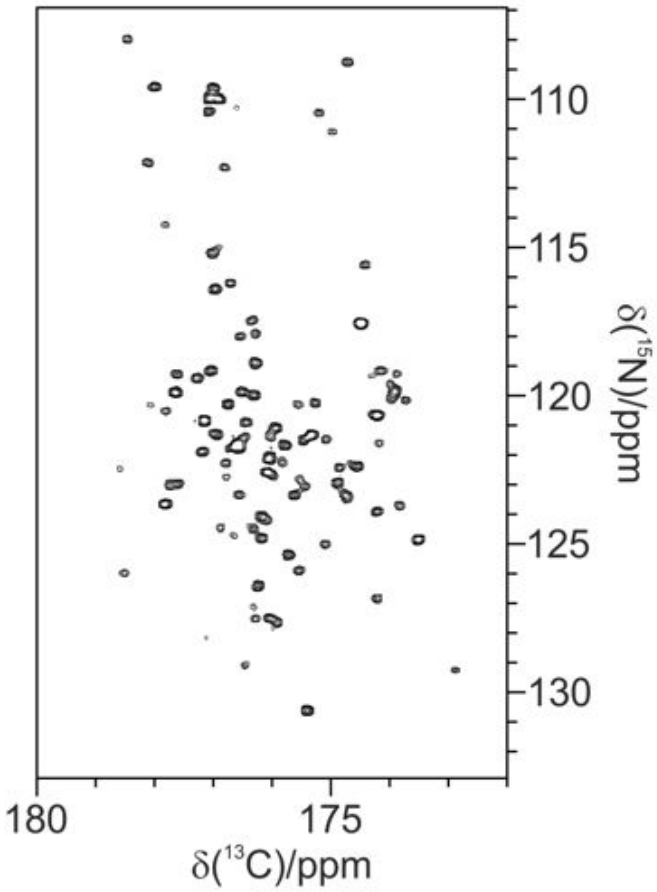


Figure 6.1. Network of backbone spin systems and the size of the ^{1}J and ^{2}J coupling constants that are frequently used for magnetization transfer in ^{13}C, ^{15}N labelled proteins.

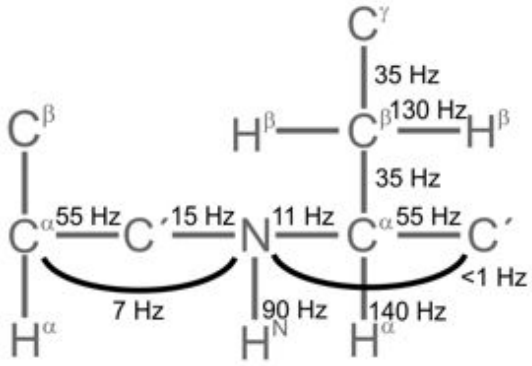


Figure 6.2. Sequential walk through the NMR spectra for the sequence-specific assignment is illustrated on the example of 3D intra-HNCA and HN(CO)CA experiments.

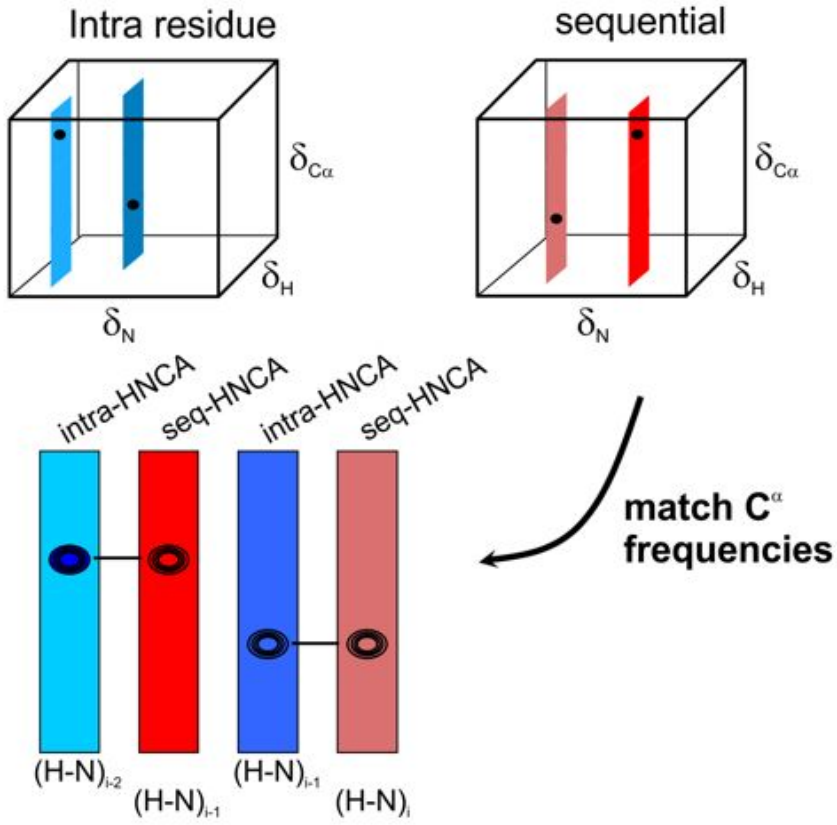


Figure 6.3. Sequence-specific assignment procedure. The fragments obtained from the sequential walk through the spectra (**a**) are combined with the primary sequence and sequence-specific assignment is achieved (**b**).

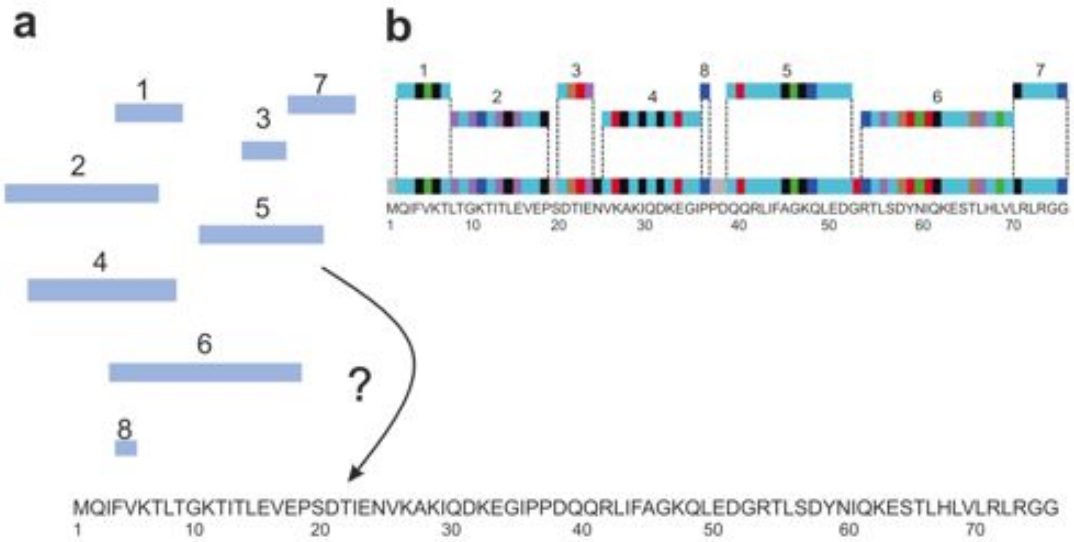


Figure 6.4. Schematic representation of the $^{1}H^{N}$ detected 3D experiments used for the sequence-specific assignment of IDPs.

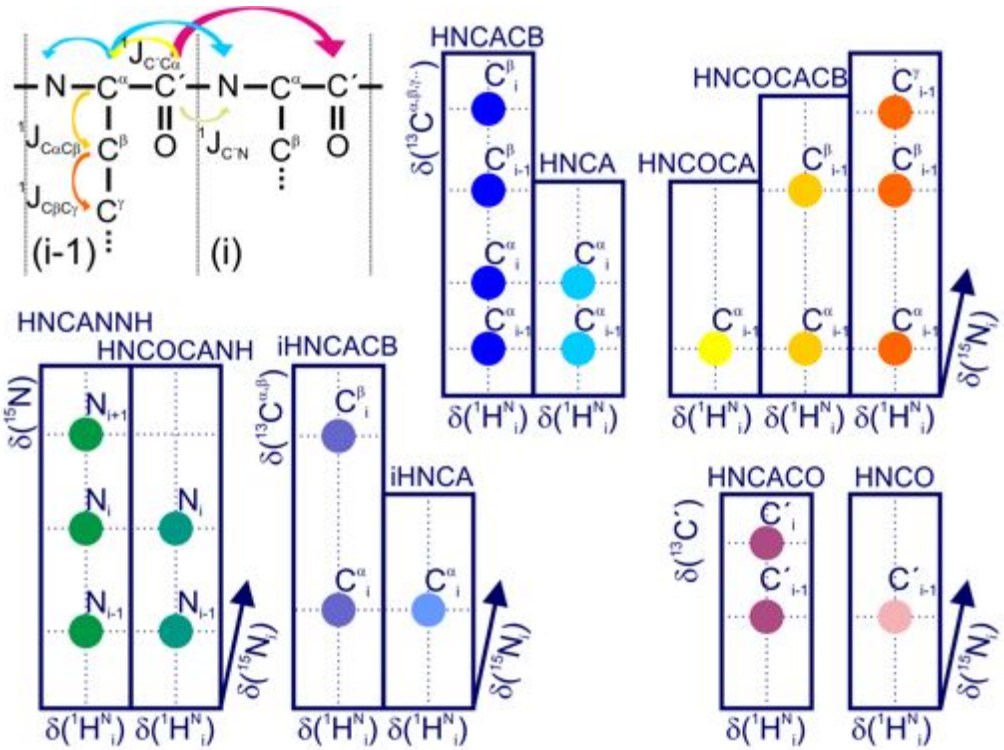


Figure 6.5. Polarization transfer pathway of HNCANNH (**a**) and HN(CO)CANH (**b**) experiments.

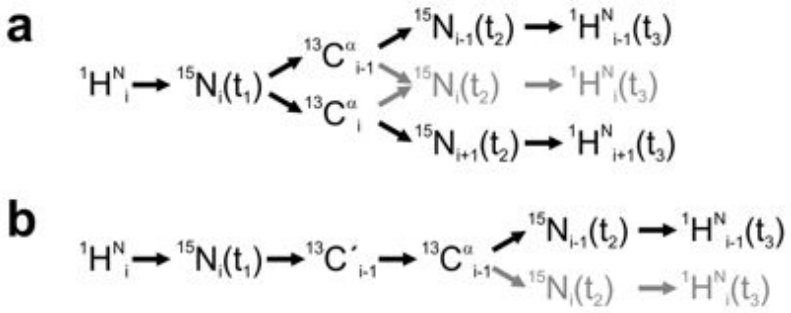


Figure 6.6. Schematic representation of the ^{13}C' detected 3D experiments used for the sequence-specific assignment of IDPs.

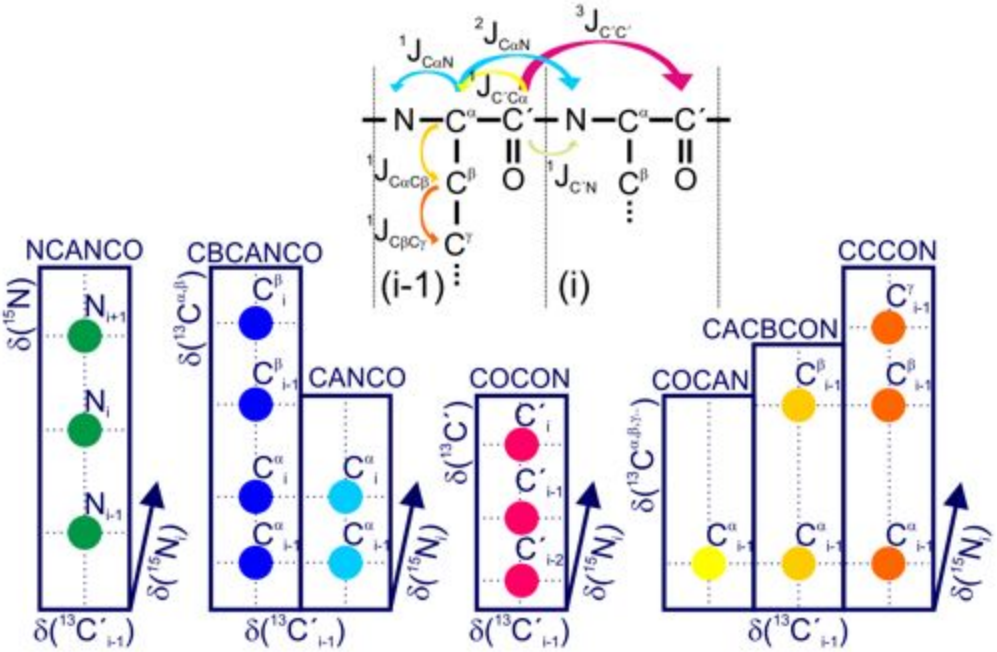


Figure 6.7. Schematic representation of the $^{1}H^{\alpha}$ detected 3D experiments used for the sequence-specific assignment of IDPs.

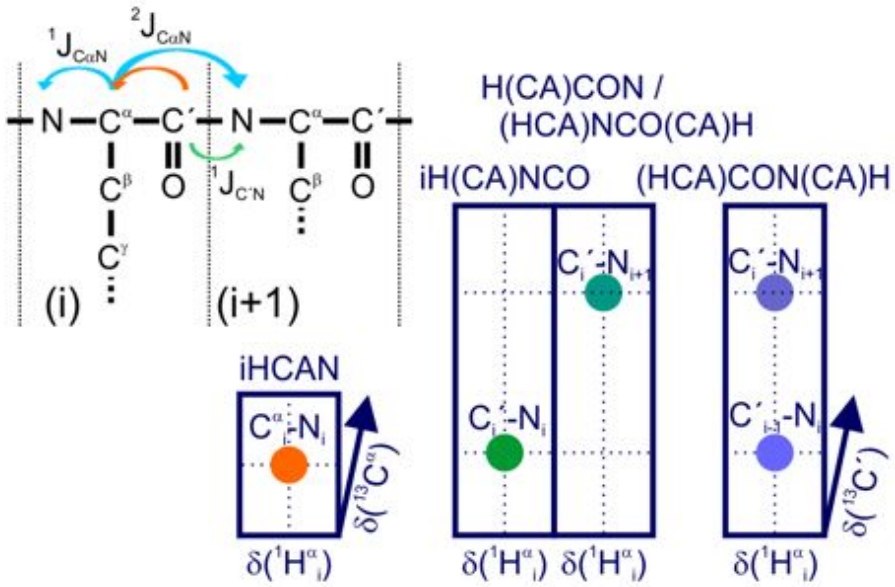


Figure 6.8. Amino acid selective experiments to simplify crowded spectra. **a** A schematic representation of a 2D CON spectrum and three 2D (HCA)CON spectra, in which three different amino acids (types A, B and C) are selected. **b** CAS-NMR spectra of Cox 17. From left to right: reference 2D CON spectrum, and (CA)CON spectra from experiments selecting Ala, Ser and Gly of Cox 17. The experiments were acquired on Cox 17 protein (1.8 mM sample in 20 mM phosphate buffer, pH 7, 0.25 mM EDTA, 20 mM DTT, at 307 K) at 700 MHz Bruker AVANACE spectrometer equipped with CPTXO probe [151].

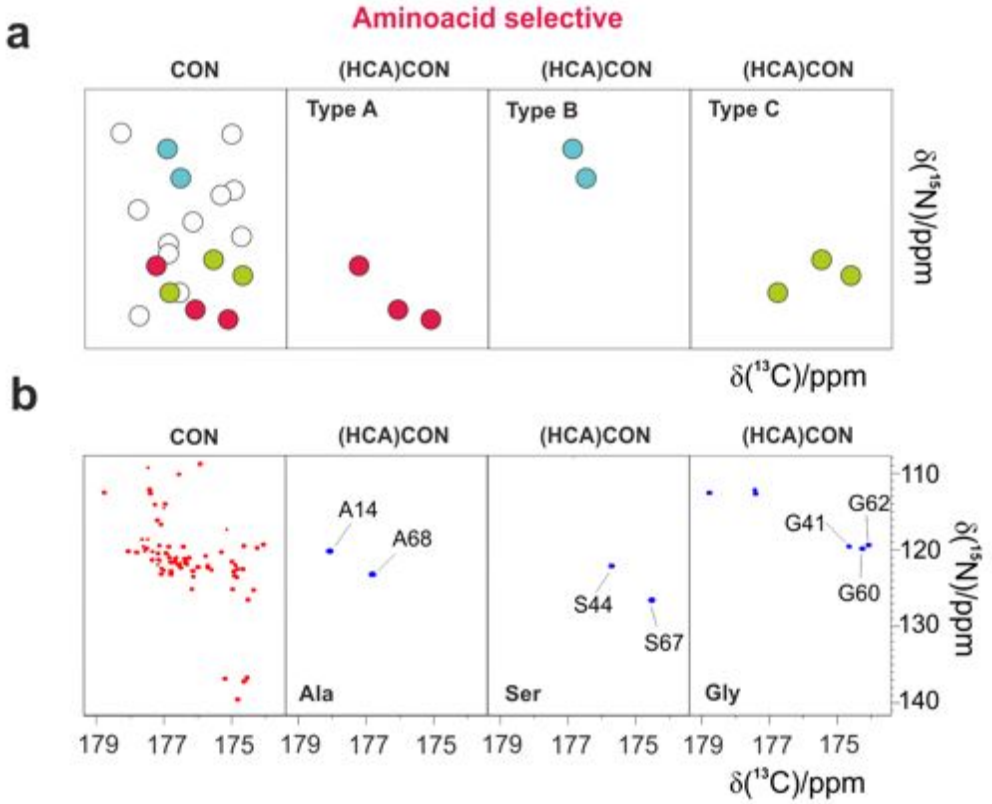


Figure 6.9. iHADAMAC (**a**) and HADAMAC (**b**) spectra acquired at 800 MHz spectrometer at 298 K on a 90 μM sample of the NS5A fragment. 100 complex points were recorded in the ^{15}N indirect dimension for a spectral width of 2000 Hz using semi-CT editing for both spectra. The total acquisition times were set to 12 h (iHADAMAC) and 2 h (sHADAMAC). The seven subspectra corresponding to the different amino acid type classes are color-coded and superposed on the same graph.

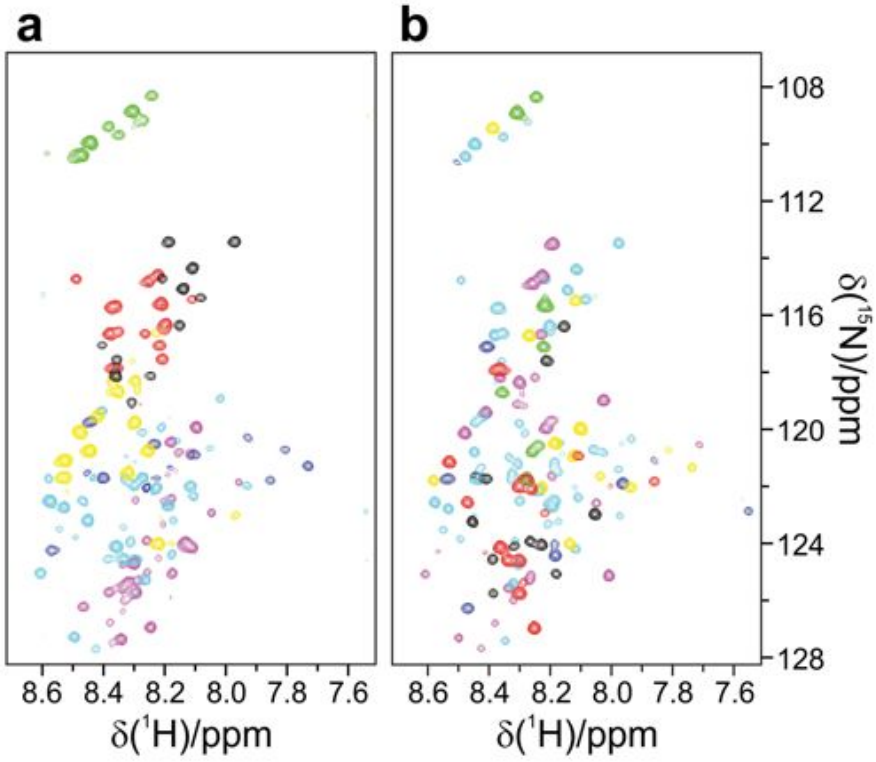


Figure 7.1. A possible way to visualize high-multidimensional NMR spectra. A schematic illustration of the progressive steps towards experiments of higher dimensionality is reported from the 1D to the 5D case. Spectra up to three dimensions are visualized in a standard way, whereas 4D and 5D spectra are presented as processed by SMFT. The algorithm exploits the prior knowledge of the correlations provided by a lower dimensionality spectrum in order to extract from a related multidimensional spectrum only the additional information, reducing thus the number of dimensions that have to be inspected. So, for example, a 4D spectrum can be thought as a 2D basis spectrum, which shares two dimensions with the 4D spectrum (δ_i and δ_j), in which each peak is associated to another 2D spectrum containing the two further dimensions (δ_l and δ_m). In this way, instead of the full spectrum, only a series of cross-sections are effectively computed, in equal number to the peaks detected in the basis spectrum. Similarly, a 5D spectrum can be analyzed as a series of two-dimensional spectra (δ_l and δ_m), each one correlated to a given peak of the related 3D basis spectrum (δ_i, δ_j and δ_k).

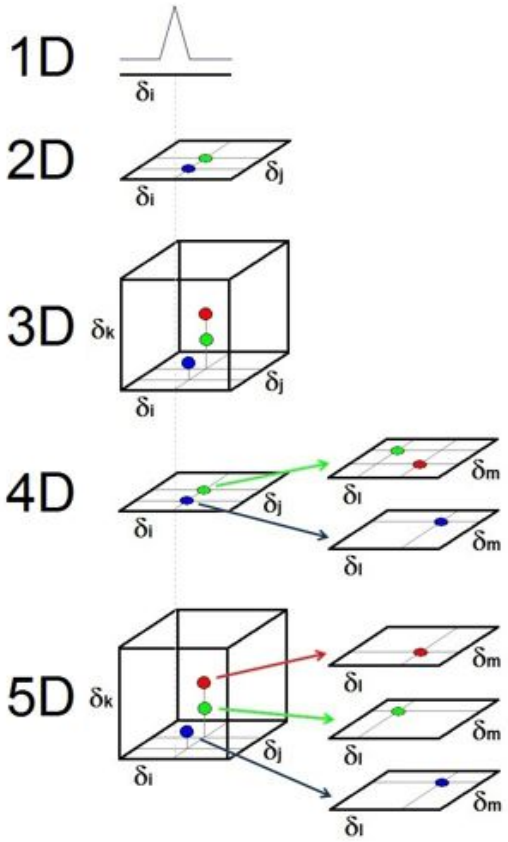


Figure 7.2. Different examples of the polarization transfer pathways of 4D and 5D experiments

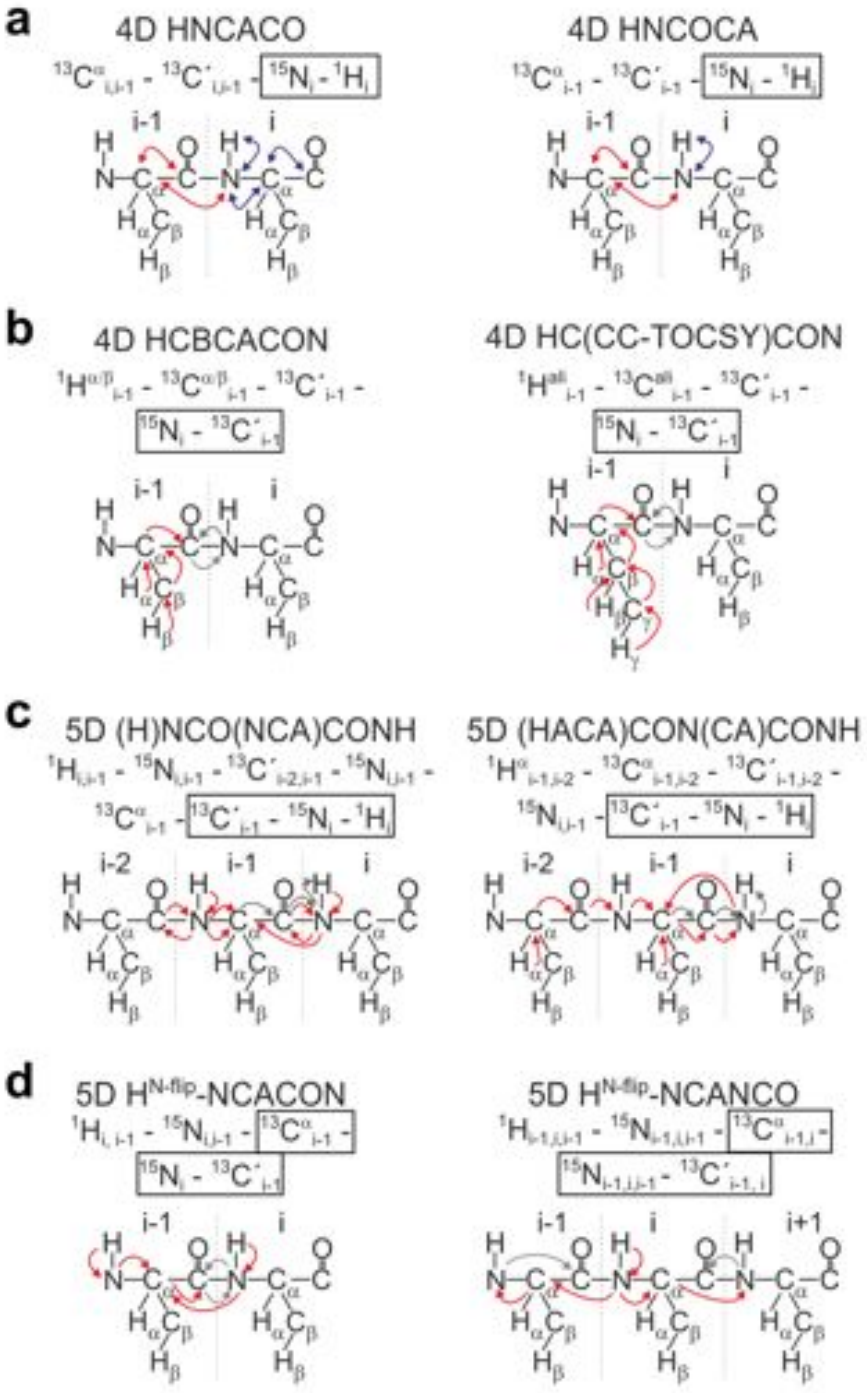
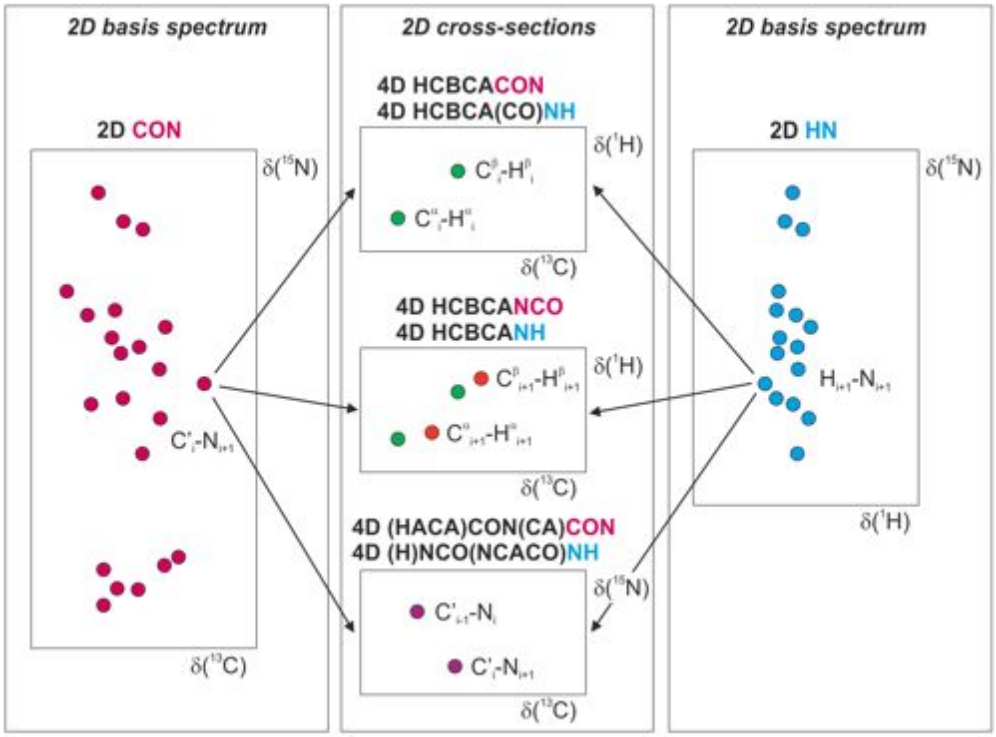


Figure 7.3. A schematic illustration of a series of ^{13}C and $^{1}H^{N}$ detected 4D spectra processed with SMFT. For each couple of experiments, the information provided is the same, except for the direct dimension. Since the 4D experiments share C'_{i}-N_{i+1} or H^{N}_{i+1}-N_{i+1} frequencies, the 2D CON (reported on the left) or 2D HN (reported on the right) can be respectively used as basis spectra to collect these common frequencies. Then, the new information content of the various 4D experiments can be easily retrieved inspecting a series of 2D cross-sections, reported in the middle part of the illustration, where the correlation provided by each experiment are shown.



References

[1] Purcell E, Torrey H, Pound R. Resonance Absorption by Nuclear Magnetic Moments in a Solid. *Phys Rev.* 1946;69(1-2):37–38.

[2] Bloch F. Nuclear induction. *Phys Rev.* 1946;70(7-8):460–474.

[3] Hoult DI, Richards RE. The signal-to-noise ratio of the nuclear magnetic resonance experiment. *J Magn Reson.* 1976;24(1):71–85.

[4] Kovacs H, Moskau D, Spraul M. Cryogenically cooled probes—a leap in NMR technology. *Progress in Nuclear Magnetic Resonance Spectroscopy.* 2005;46(2):131–155.

[5] Schanda P, Forge V, Brutscher B. HET-SOFAST NMR for fast detection of structural compactness and heterogeneity along polypeptide chains. *Magn Reson Chem.* 2006;44(S1):S177–S184.

[6] Gil S, Hošek T, Solyom Z, Kümmerle R, Brutscher B, Pierattelli R, Felli IC. NMR Spectroscopic Studies of Intrinsically Disordered Proteins at Near-Physiological Conditions. *Angew Chem Int Ed.* 2013;52(45):11808–11812.

[7] Spera S, Bax A. Empirical correlation between protein backbone conformation and C. alpha. and C. beta. 13C nuclear magnetic resonance chemical shifts. *J Am Chem Soc.* 1991;113(14):5490–5492.

[8] Wishart DS, Sykes BD, Richards FM. Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *Journal of Molecular Biology.* 1991;222(2):311–333.

[9] Wishart DS, Sykes BD, Richards FM. The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry.* 1992;31(6):1647–1651.

[10] Wishart DS, Sykes BD. The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. *J Biomol NMR.* 1994;4(2):171–180.

[11] Schwarzinger S, Kroon GJA, Foss TR, Chung J, Wright PE, Dyson HJ. Sequence-Dependent Correction of Random Coil NMR Chemical Shifts. *J Am Chem Soc.* 2001;123(13):2970–2978.

[12] Kjaergaard M, Brander S, Poulsen FM. Random coil chemical shift for intrinsically disordered proteins: effects of temperature and pH. *J Biomol NMR*. 2011;49(2):139–149.

[13] Kjaergaard M, Poulsen FM. Disordered proteins studied by chemical shifts. *Progress in Nuclear Magnetic Resonance Spectroscopy*. 2012;60(C):42–51.

[14] Tamiola K, Acar B, Mulder FAA. Sequence-Specific Random Coil Chemical Shifts of Intrinsically Disordered Proteins. *J Am Chem Soc*. 2010;132(51):18000–18003.

[15] Zhang H, Neal S, Wishart DS. RefDB: a database of uniformly referenced protein chemical shifts. *J Biomol NMR*. 2003;25(3):173–195.

[16] Wishart DS, Bigam CG, Holm A, Hodges RS, Sykes BD. 1H, 13C and 15N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *J Biomol NMR*. 1995;5(1):67–81.

[17] Kjaergaard M, Poulsen FM. Sequence correction of random coil chemical shifts: correlation between neighbor correction factors and changes in the Ramachandran distribution. *J Biomol NMR*. 2011;50(2):157–165.

[18] Karplus M. Contact Electron-Spin Coupling of Nuclear Magnetic Moments. *J Chem Phys*. 1959;30(1):11–15.

[19] Billeter M, Neri D, Otting G, Qian YQ, Wüthrich K. Precise vicinal coupling constants3JHNα in proteins from nonlinear fits of J-modulated [15N,1H]-COSY experiments. *J Biomol NMR*. 1992;2(3):257–274.

[20] Vuister GW, Bax A. Quantitative J correlation: a new approach for measuring homonuclear three-bond J(HNH.alpha.) coupling constants in 15N-enriched proteins. *J Am Chem Soc*. 1993;115(17):7772–7777.

[21] Case DA. Interpretation of chemical shifts and coupling constants in macromolecules. *Current Opinion in Structural Biology*. 2000;10(2):197–203.

[22] Otten R, Wood K, Mulder FAA. Comprehensive determination of 3JHNHα for unfolded proteins using 13C-resolved spin-echo difference spectroscopy. *J Biomol NMR*. 2009;45(4):343–349.

[23] Tolman JR, Flanagan JM, Kennedy MA, Prestegard JH. Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution. *Proc Natl Acad Sci USA*. 1995;92(20):9279–9283.

[24] Tjandra N, Grzesiek S, Bax A. Magnetic Field Dependence of NitrogenProton J Splittings in 15N-Enriched Human Ubiquitin Resulting from Relaxation Interference and Residual Dipolar Coupling. *Journal of the American Chemical ...*. 1996; 118(26):6264–6272.

[25] Banci L, Bertini I, Huber JG, Luchinat C, Rosato A. Partial Orientation of Oxidized and Reduced Cytochrome b5at High Magnetic Fields: Magnetic Susceptibility Anisotropy Contributions and Consequences for Protein Solution Structure Determination. *J Am Chem Soc*. 1998;120(49):12903–12909.

[26] Bax A, Grishaev A. Weak alignment NMR: a hawk-eyed view of biomolecular structure. *Current Opinion in Structural Biology*. 2004;15(5):563–570.

[27] Peng JW, Wagner G. Investigation of protein motions via relaxation measurements. *Methods Enzymol*. 1994;239:563–596.

[28] Barbato G, Ikura M, Kay LE, Pastor RW. Backbone Dynamics of Calmodulin Studied by N-15 Relaxation Using Inverse Detected 2-Dimensional Nmr-Spectroscopy - the Central Helix Is Flexible. *Biochemistry*. 1992;31(23):5269–5278.

[29] Cavanagh J, Fairbrother WJ, Palmer AG III, Skelton NJ, Rance M. *Protein NMR Spectroscopy, Second Edition: Principles and Practice*. San Diego: Academic Press, 2nd ed. 2006.

[30] Neuhaus D, Williamson MP. *The Nuclear Overhauser Effect in Structural and Conformational Analysis*. New York: Wiley. 1989.

[31] Bloch F. Dynamical Theory of Nuclear Induction. II. *Phys Rev*. 1956;102(1):104–135.

[32] Wangsness R, Bloch F. The Dynamical Theory of Nuclear Induction. *Phys Rev*. 1953;89(4):728–739.

[33] Redfield AG. On the theory of relaxation processes. *IBM J Res Develop*. 1957; 1:19–31.

[34] Lipari G, Szabo A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J Am Chem Soc*. 1982;104(17):4546–4559.

[35] Kay LE, Torchia DA, Bax A. Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry*. 1989;28(23):8972–8979.

[36] Farrow NA, Zhang O, Szabo A, Torchia DA, Kay LE. Spectral density function mapping using 15N relaxation data exclusively. *J Biomol NMR*. 1995;6(2):153–162.

[37] Kadeřávek P, Zapletal V, Rabatinová A, Krásný L, Sklenář V, Žídek L. Spectral density mapping protocols for analysis of molecular motions in disordered proteins. *J Biomol NMR*. 2014;58(3):193–207.

[38] Tompa P. Intrinsically unstructured proteins. *Trends in Biochemical Sciences*. 2002; 27(10):527–533.

[39] Tycko R. Solid-state NMR as a probe of amyloid structure. *Protein Pept Lett*. 2006; 13(3):229–234.

[40] Chimon S, Shaibat MA, Jones CR, Calero DC, Aizezi B, Ishii Y. Evidence of fibril-like β-sheet structures in a neurotoxic amyloid intermediate of Alzheimer's β-amyloid. *Nat Struct Mol Biol*. 2007;14(12):1157–1164.

[41] Bertini I, Luchinat C, Parigi G, Ravera E, Reif B, Turano P. Solid-state NMR of proteins sedimented by ultracentrifugation. *Proc Natl Acad Sci USA*. 2011; 108(26):10396–10399.

[42] Theillet FX, Kalmar L, Tompa P, Han KH, Selenko P, Dunker AK, Daughdrill GW, Uversky VN. The alphabet of intrinsic disorder. *Intrinsically Disord Proteins*. 2014; 1(1):e24360.

[43] Zhang YZ. Protein and Peptide Structure and Interactions Studied by Hydrogen Exchange and NMR. University of Pennsylvania. 1995.

[44] Bai Y, Milne JS, Mayne L, Englander SW. Primary structure effects on peptide group hydrogen exchange. *Proteins*. 1993;17(1):75–86.

[45] Sattler M, Schleucher J, Griesinger C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Progress in Nuclear Magnetic Resonance Spectroscopy*. 1999;34(2):93–158.

[46] Lescop E, Schanda P, Brutscher B. A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *Journal of Magnetic Resonance*. 2007;187(1):163–169.

[47] Felli IC, Pierattelli R. Novel methods based on 13C detection to study intrinsically disordered proteins. *Journal of Magnetic Resonance*. 2014;241(C):115–125.

[48] Takeuchi K, Heffron G, Sun ZYJ, Frueh DP, Wagner G. Nitrogen-detected CAN and CON experiments as alternative experiments for main chain NMR resonance assignments. *J Biomol NMR*. 2010;47(4):271–282.

[49] Gal M, Edmonds KA, Milbradt AG, Takeuchi K, Wagner G. Speeding up direct (15)N detection: hCaN 2D NMR experiment. *J Biomol NMR*. 2011;51(4):497–504.

[50] Vasos PR, Hall JB, Kümmerle R, Fushman D. Measurement of 15N relaxation in deuterated amide groups in proteins using direct nitrogen detection. *J Biomol NMR*. 2006;36(1):27–36.

[51] Favier A, Brutscher B. Recovering lost magnetization: polarization enhancement in biomolecular NMR. *J Biomol NMR*. 2011;49(1):9–15.

[52] Bermel W, Bertini I, Felli IC, Kümmerle R, Pierattelli R. Novel 13C direct detection experiments, including extension to the third dimension, to perform the complete assignment of proteins. *Journal of Magnetic Resonance*. 2006;178(1):56–64.

[53] Selenko P, Frueh DP, Elsaesser SJ, Haas W, Gygi SP, Wagner G. In situ observation of protein phosphorylation by high-resolution NMR spectroscopy. *Nat Struct Mol Biol*. 2008;15(3):321–329.

[54] Felli IC, Gonnelli L, Pierattelli R. In-cell C NMR spectroscopy for the study of intrinsically disordered proteins. *Nat Protoc*. 2014;9(9):2005–2016.

[55] Selenko P, Wagner G. Looking into live cells with in-cell NMR spectroscopy. *J Struct Biol*. 2007;158(2):244–253.

[56] Serber Z, Selenko P, Hänsel R, Reckel S, Löhr F, Ferrell JE, Wagner G, Dötsch V. Investigating macromolecules inside cultured and injected cells by in-cell NMR spectroscopy. *Nat Protoc.* 2007;1(6):2701–2709.

[57] Ernst RR, Bodenhausen G, Wokaun A. *Principles of Nuclear Magnetic Resonance in One and Two Dimensions.* Oxford University Press. 1990.

[58] Morris GA, Freeman R. Enhancement of nuclear magnetic resonance signals by polarization transfer. *J Am Chem Soc.* 1979;101(3):760–762.

[59] Brutscher B, Morelle N, Cordier F, Marion D. Determination of an Initial Set of NOE-Derived Distance Constraints for the Structure Determination of15N/13C-Labeled Proteins. *J Magn Reson B.* 1995;109(2):238–242.

[60] Szyperski T, Wider G, Bushweller JH, Wüthrich K. Reduced dimensionality in triple-resonance NMR experiments. *J Am Chem Soc.* 1993;115(20):9307–9308.

[61] Kupče E, Freeman R. Projection-reconstruction of three-dimensional NMR spectra. *J Am Chem Soc.* 2003;125(46):13958–13959.

[62] Luan T, Jaravine V, Yee A, Arrowsmith CH, Orekhov VY. Optimization of resolution and sensitivity of 4D NOESY using multi-dimensional decomposition. *J Biomol NMR.* 2005;33(1):1–14.

[63] Tugarinov V, Kay LE, Ibraghimov I, Orekhov VY. High-resolution four-dimensional 1H-13C NOE spectroscopy using methyl-TROSY, sparse data acquisition, and multidimensional decomposition. *J Am Chem Soc.* 2005;127(8):2767–2775.

[64] Hoch JC, Stern A. *NMR Data Processing.* New York: Wiley-Interscience. 1996.

[65] Kazimierczuk K, Orekhov VY. Accelerated NMR Spectroscopy by Using Compressed Sensing. *Angew Chem Int Ed.* 2011;50(24):5556–5559.

[66] Holland DJ, Bostock MJ, Gladden LF, Nietlispach D. Fast multidimensional NMR spectroscopy using compressed sensing. *Angew Chem Int Ed.* 2011;50(29):6548–6551.

[67] Kazimierczuk K, Stanek J, Zawadzka-Kazimierczuk A, Koźmiński W. Random sampling in multidimensional NMR spectroscopy. *Progress in Nuclear Magnetic Resonance Spectroscopy.* 2010;57(4):420–434.

[68] Matsuki Y, Eddy MT, Herzfeld J. Spectroscopy by integration of frequency and time domain information for fast acquisition of high-resolution dark spectra. *J Am Chem Soc*. 2009;131(13):4648–4656.

[69] Levitt MH, Freeman R, Frenkiel T. Broadband heteronuclear decoupling. *J Magn Reson*. 1982;47(2):328–330.

[70] Shaka AJ, Lee CJ, Pines A. Iterative schemes for bilinear operators; application to spin decoupling. *J Magn Reson*. 1988;77(2):274–293.

[71] Shaka AJ, Keeler J, Freeman R. Evaluation of a new broadband decoupling sequence: WALTZ-16. *J Magn Reson*. 1983;53(2):313–340.

[72] Shaka AJ, Keeler J, Frenkiel T, Freeman R. An improved sequence for broadband decoupling: WALTZ-16. *J Magn Reson*. 1983;52(2):335–338.

[73] Shaka AJ, Barker PB, Freeman R. Computer-optimized decoupling scheme for wideband applications and low-level operation. *J Magn Reson (1969-1992)*. 1985; 64(3):547–552.

[74] Pervushin K, Riek R, Wider G, Wüthrich K. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci USA*. 1997;94(23):12366–12371.

[75] Schanda P. Fast-pulsing longitudinal relaxation optimized techniques: Enriching the toolbox of fast biomolecular NMR spectroscopy. *Progress in Nuclear Magnetic Resonance Spectroscopy*. 2009;55(3):238–265.

[76] Solomon I. Relaxation Processes in a System of Two Spins. *Phys Rev*. 1955;99:559–565.

[77] McConnell HM. Reaction Rates by Nuclear Magnetic Resonance. *J Chem Phys*. 1958;28(3):430–431.

[78] Solyom Z, Schwarten M, Geist L, Konrat R, Willbold D, Brutscher B. BEST-TROSY experiments for time-efficient sequential resonance assignment of large disordered proteins. *J Biomol NMR*. 2013;55(4):311–321.

[79] Schanda P, Brutscher B. Very Fast Two-Dimensional NMR Spectroscopy for Real-Time Investigation of Dynamic Events in Proteins on the Time Scale of Seconds. *J Am Chem Soc.* 2005;127(22):8014–8015.

[80] Schanda P, Van Melckebeke H, Brutscher B. Speeding up three-dimensional protein NMR experiments to a few minutes. *J Am Chem Soc.* 2006;128(28):9042–9043.

[81] Felli IC, Piai A, Pierattelli R. Recent Advances in Solution NMR Studies: 13C Direct Detection for Biomolecular NMR Applications. *Ann Rep NMR Spectroscop.* 2013;(6):359–418.

[82] Emsley L, Bodenhausen G. Phase shifts induced by transient Bloch-Siegert effects in NMR. *Chem Phys Lett.* 1990;168(3):297–303.

[83] Waugh JS. Theory of broadband spin decoupling. *J Magn Reson.* 1982;50(1):30–49.

[84] Ying J, Li F, Lee JH, Bax A. 13Cα decoupling during direct observation of carbonyl resonances in solution NMR of isotopically enriched proteins. *J Biomol NMR.* 2014; 60(1):15–21.

[85] Shimba N, Stern AS, Craik CS, Hoch JC, Dötsch V. Elimination of 13Calpha splitting in protein NMR spectra by deconvolution with maximum entropy reconstruction. *J Am Chem Soc.* 2003;125(9):2382–2383.

[86] Felli IC, Pierattelli R. Spin-State-Selective Methods in Solution- and Solid-State Biomolecular 13C NMR. *Progress in Nuclear Magnetic Resonance Spectroscopy.* 2014;in press.

[87] Duma L, Hediger S, Brutscher B, Böckmann A, Emsley L. Resolution enhancement in multidimensional solid-state NMR spectroscopy of proteins using spin-state selection. *J Am Chem Soc.* 2003;125(39):11816–11817.

[88] Duma L, Hediger S, Lesage A, Emsley L. Spin-state selection in solid-state NMR. *J Magn Reson.* 2003;164(1):187–195.

[89] Bertini I, Felli IC, Kümmerle R, Luchinat C, Pierattelli R. (13)C-(13)C NOESY: a constructive use of (13)C-(13)C spin-diffusion. *J Biomol NMR.* 2004;30(3):245–251.

[90] Bermel W, Felli IC, Kümmerle R, Pierattelli R. 13C Direct-detection biomolecular NMR. *Concepts Magn Reson.* 2008;32A(3):183–200.

[91] Kern T, Schanda P, Brutscher B. Sensitivity-enhanced IPAP-SOFAST-HMQC for fast-pulsing 2D NMR with reduced radiofrequency load. *J Magn Reson.* 2008; 190(2):333–338.

[92] Bermel W, Bertini I, Csizmok V, Felli IC, Pierattelli R, Tompa P. H-start for exclusively heteronuclear NMR spectroscopy: The case of intrinsically disordered proteins. *Journal of Magnetic Resonance.* 2009;198(2):275–281.

[93] Shimba N, Kovacs H, Stern AS, Nomura AM, Shimada I, Hoch JC, Craik CS, Dötsch V. Optimization of 13C direct detection NMR methods. *J Biomol NMR.* 2004; 30(2):175–179.

[94] O'Hare B, Benesi AJ, Showalter SA. Incorporating1H chemical shift determination into13C-direct detected spectroscopy of intrinsically disordered proteins in solution. *Journal of Magnetic Resonance.* 2009;200(2):354–358.

[95] Bermel W, Bertini I, Felli IC, Pierattelli R. Speeding Up 13C Direct Detection Biomolecular NMR Spectroscopy. *J Am Chem Soc.* 2009;131(42):15339–15345.

[96] Bertini I, Felli IC, Gonnelli L, Kumar V M V, Pierattelli R. High-Resolution Characterization of Intrinsic Disorder in Proteins: Expanding the Suite of 13C-Detected NMR Spectroscopy Experiments to Determine Key Observables. *ChemBioChem.* 2011;12(15):2347–2352.

[97] Nováček J, Zawadzka-Kazimierczuk A, Papoušková V, Žídek L, Šanderová H, Krásný L, Koźmiński W, Sklenář V. 5D 13C-detected experiments for backbone assignment of unstructured proteins with a very low signal dispersion. *J Biomol NMR.* 2011;50(1):1–11.

[98] Bermel W, Bertini I, Duma L, Felli IC, Emsley L, Pierattelli R, Vasos PR. Complete Assignment of Heteronuclear Protein Resonances by Protonless NMR Spectroscopy. *Angew Chem Int Ed.* 2005;44(20):3089–3092.

[99] Wüthrich K. *NMR of Proteins and Nucleic Acids.* Wiley-Interscience. 1986.

[100] Dyson HJ, Wright PE. Nuclear magnetic resonance methods for elucidation of structure and dynamics in disordered states. *Methods Enzymol.* 2001;339:258–270.

[101] Eliezer D. Biophysical characterization of intrinsically disordered proteins. *Current Opinion in Structural Biology.* 2009;19(1):23–30.

[102] Ikura M, Kay LE, Bax A. A novel approach for sequential assignment of 1H, 13C, and 15N spectra of proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. *Biochemistry*. 1990;29(19):4659–4667.

[103] Kay LE, Ikura M, Tschudin R, Bax A. Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *J Magn Reson (1969-1992)*. 1990; 89(3):496–514.

[104] Grzesiek S, Bax A. Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. *J Magn Reson*. 1992;96:432–440.

[105] Schleucher J, Sattler M, Griesinger C. Coherence Selection by Gradients without Signal Attenuation: Application to the Three-Dimensional HNCO Experiment. *Angew Chem Int Ed*. 1993;32(10):1489–1491.

[106] Bax A, Ikura M. An efficient 3D NMR technique for correlating the proton and 15N backbone amide resonances with the alpha-carbon of the preceding residue in uniformly 15N/13C enriched proteins. *J Biomol NMR*. 1991;1(1):99–104.

[107] Grzesiek S, Bax A. Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J Am Chem Soc*. 1992; 114(16):6291–6293.

[108] Yamazaki T, Lee W, Arrowsmith CH, Muhandiram DR, Kay LE. A Suite of Triple Resonance NMR Experiments for the Backbone Assignment of 15N, 13C, 2H Labeled Proteins with High Sensitivity. *J Am Chem Soc*. 1994;116(26):11655–11666.

[109] Wittekind M, Mueller L. HNCACB, a High-Sensitivity 3D NMR Experiment to Correlate Amide-Proton and Nitrogen Resonances with the Alpha- and Beta-Carbon Resonances in Proteins. *J Magn Reson B*. 1993;101(2):201–205.

[110] Muhandiram DR, Kay LE. Gradient-Enhanced Triple-Resonance Three-Dimensional NMR Experiments with Improved Sensitivity. *J Magn Reson B*. 1994;103(3):203–216.

[111] Clubb RT, Thanabal V, Wagner G. A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue1HN,15N, and13C chemical shifts in15N-13C-labelled proteins. *J Magn Reson*. 1992;97(1):213–217.

[112] Kay LE, Xu GY, Yamazaki T. Enhanced-Sensitivity Triple-Resonance Spectroscopy with Minimal H2O Saturation. *J Magn Reson A*. 1994;109(1):129–133.

[113] Briand L, Lescop E, Bézirard V, Birlirakis N, Huet JC, Henry C, Guittet E, Pernollet JC. Isotopic Double-Labeling of Two Honeybee Odorant-Binding Proteins Secreted by the Methylotrophic YeastPichia pastoris. *Protein Expression and Purification*. 2001;23(1):167–174.

[114] Nietlispach D, Ito Y, Laue ED. A novel approach for the sequential backbone assignment of larger proteins: selective intra-HNCA and DQ-HNCA. *J Am Chem Soc*. 2002;124(37):11199–11207.

[115] Brutscher B. Intraresidue HNCA and COHNCA Experiments for Protein Backbone Resonance Assignment. *Journal of Magnetic Resonance*. 2002;156(1):155–159.

[116] Nietlispach D. A selective intra-HN (CA) CO experiment for the backbone assignment of deuterated proteins. *J Biomol NMR*. 2004;28(2):131–136.

[117] Gil-Caballero S, Favier A, Brutscher B. HNCA plus, HNCO plus, and HNCACB plus experiments: improved performance by simultaneous detection of orthogonal coherence transfer pathways. *J Biomol NMR*. 2014;60(1):1–9.

[118] Grzesiek S, Anglister J, Ren H, Bax A. 13C line narrowing by 2H decoupling in 2H/13C/15N-enriched proteins. Application to triple resonance 4D J connectivity of sequential amides. *J Am Chem Soc*. 1993;115(10):4369–4370.

[119] Weisemann R, Rüterjans H, Bermel W. 3D Triple-resonance NMR techniques for the sequential assignment of NH and 15N resonances in 15N- and 13C-labelled proteins. *J Biomol NMR*. 1993;3(1):113–120.

[120] Bracken C, Palmer AG III, Cavanagh J. (H)N(COCA)NH and HN(COCA)NH experiments for 1H-15N backbone assignments in 13C/15N-labeled proteins. *J Biomol NMR*. 1997;9(1):94–100.

[121] Panchal SC, Bhavesh NS, Hosur RV. Improved 3D triple resonance experiments, HNN and HN(C)N, for HN and 15N sequential correlations in (13C, 15N) labeled proteins: application to unfolded proteins. *J Biomol NMR*. 2001;20(2):135–147.

[122] Kumar D, Hosur RV. hNCOcanH pulse sequence and a robust protocol for rapid and unambiguous assignment of backbone (H-1(N), N-15 and C-13 ') resonances in N-15/C-13-labeled proteins. *Magn Reson Chem*. 2011;49(9):575–583.

[123] Montelione GT, Lyons BA, Emerson SD, Tashiro M. An efficient triple resonance experiment using carbon-13 isotropic mixing for determining sequence-specific resonance assignments of isotopically-enriched proteins. *J Am Chem Soc.* 1992; 114(27):10974–10975.

[124] Logan TM, Olejniczak ET, Xu R, Fesik SW. Side chain and backbone assignments in isotopically labeled proteins from two heteronuclear triple resonance experiments. *FEBS Letters.* 1992;314(3):413–418.

[125] Logan TM, Olejniczak ET, Xu RX, Fesik SW. A general method for assigning NMR spectra of denatured proteins using 3D HC(CO)NH-TOCSY triple resonance experiments. *J Biomol NMR.* 1993;3(2):225–231.

[126] Grzesiek S, Anglister J, Bax A. Correlation of Backbone Amide and Aliphatic Side-Chain Resonances in13C/15N-Enriched Proteins by Isotropic Mixing of13C Magnetization. *J Magn Reson B.* 1993;101(1):114–119.

[127] Gardner KH, Konrat R, Rosen MK, Kay LE. An (H)C(CO)NH-TOCSY pulse scheme for sequential assignment of protonated methyl groups in otherwise deuterated (15)N, (13)C-labeled proteins. *J Biomol NMR.* 1996;8(3):351–356.

[128] Bermel W, Bertini I, Felli IC, Piccioli M, Pierattelli R. 13C-detected protonless NMR spectroscopy of proteins in solution. *Progress in Nuclear Magnetic Resonance Spectroscopy.* 2006;48(1):25–45.

[129] Bermel W, Bertini I, Felli IC, Lee YM, Luchinat C, Pierattelli R. Protonless NMR Experiments for Sequence-Specific Assignment of Backbone Nuclei in Unfolded Proteins. *J Am Chem Soc.* 2006;128(12):3918–3919.

[130] Felli IC, Pierattelli R, Glaser SJ, Luy B. Relaxation-optimised Hartmann–Hahn transfer using a specifically Tailored MOCCA-XY16 mixing sequence for carbonyl–carbonyl correlation spectroscopy in 13C direct detection NMR experiments. *J Biomol NMR.* 2009;43(3):187–196.

[131] Kanelis V, Donaldson L, Muhandiram DR, Rotin D, Forman-Kay JD, Kay LE. Sequential assignment of proline-rich regions in proteins: application to modular binding domain complexes. *J Biomol NMR.* 2000;16(3):253–259.

[132] Mäntylahti S, Aitio O, Hellman M, Permi P. HA-detected experiments for the backbone assignment of intrinsically disordered proteins. *J Biomol NMR*. 2010; 47(3):171–181.

[133] Mäntylahti S, Hellman M, Permi P. Extension of the HA-detection based approach: (HCA)CON(CA)H and (HCA)NCO(CA)H experiments for the main-chain assignment of intrinsically disordered proteins. *J Biomol NMR*. 2011;49(2):99–109.

[134] McIntosh LP, Dahlquist FW. Biosynthetic incorporation of 15N and 13C for assignment and interpretation of nuclear magnetic resonance spectra of proteins. *Q Rev Biophys*. 1990;23(1):1–38.

[135] Tong KI, Yamamoto M, Tanaka T. A simple method for amino acid selective isotope labeling of recombinant proteins in E. coli. *J Biomol NMR*. 2008;42(1):59–67.

[136] Cowburn D, Shekhtman A, Xu R, Ottesen JJ, Muir TW. Segmental isotopic labeling for structural biological applications of NMR. *Methods Mol Biol*. 2004;278:47–56.

[137] Grzesiek S, Bax A. Amino acid type determination in the sequential assignment procedure of uniformly 13C/15N-enriched proteins. *J Biomol NMR*. 1993;3(2):185–204.

[138] Yamazaki T, Pascal SM, Singer AU, Forman-Kay JD, Kay LE. NMR pulse schemes for the sequence-specific assignment of arginine guanidino 15N and 1H chemical shifts in proteins. *J Am Chem Soc*. 1995;117(12):3556–3564.

[139] Rao NS, Legault P, Muhandiram DR, Greenblatt J, Battiste JL, Williamson JR, Kay LE. NMR pulse schemes for the sequential assignment of arginine side-chain H epsilon protons. *J Magn Reson B*. 1996;113(3):272–276.

[140] Pellecchia M, Wider G, Iwai H, Wüthrich K. Arginine side chain assignments in uniformly 15N-labeled proteins using the novel 2D HE(NE)HGHH experiment. *J Biomol NMR*. 1997;10(2):193–197.

[141] Wittekind M, Metzler WJ, Mueller L. Selective Correlations of Amide Groups to Glycine Alpha Protons in Proteins. *J Magn Reson B*. 1993;101(2):214–217.

[142] Olejniczak ET, Fesik SW. Two-Dimensional Nuclear Magnetic Resonance Method for Identifying the H-Alpha/C-Alpha Signals of Amino Acid Residues Preceding Proline. *J Am Chem Soc*. 1994;116(5):2215–2216.

[143] Rios CB, Feng W, Tashiro M, Shang Z, Montelione GT. Phase labeling of C-H and C-C spin-system topologies: Application in constant-time PFG-CBCA(CO)NH experiments for discriminating amino acid spin-system types. *J Biomol NMR*. 1996; 8(3):345–350.

[144] Dotsch V, Oswald RE, Wagner G. Selective identification of threonine, valine, and isoleucine sequential connectivities with a TVI-CBCACONH experiment. *J Magn Reson B*. 1996;110(3):304–308.

[145] Dotsch V, Oswald RE, Wagner G. Amino-acid-type-selective triple-resonance experiments. *J Magn Reson B*. 1996;110(1):107–111.

[146] Schubert M, Smalla M, Schmieder P, Oschkinat H. MUSIC in Triple-Resonance Experiments: Amino Acid Type-Selective 1H–15N Correlations. *J Magn Reson*. 1999;141(1):34–43.

[147] Schubert M, Oschkinat H, Schmieder P. MUSIC, Selective Pulses, and Tuned Delays: Amino Acid Type-Selective 1H–15N Correlations, II. *Journal of Magnetic Resonance*. 2001;148(1):61–72.

[148] Schubert M, Oschkinat H, Schmieder P. MUSIC and Aromatic Residues: Amino Acid Type-Selective 1H–15N Correlations, III. *Journal of Magnetic Resonance*. 2001; 153(2):186–192.

[149] Pantoja-Uceda D, Santoro J. Amino acid type identification in NMR spectra of proteins via β- and γ-carbon edited experiments. *Journal of Magnetic Resonance*. 2008;195(2):187–195.

[150] Löhr F, Rüterjans H. H2NCO–E.COSY, a Simple Method for the Stereospecific Assignment of Side-Chain Amide Protons in Proteins. *Journal of Magnetic Resonance*. 1997;124(1):255–258.

[151] Bermel W, Bertini I, Chill J, Felli IC, Haba N, Kumar V M V, Pierattelli R. Exclusively Heteronuclear 13C-Detected Amino-Acid-Selective NMR Experiments for the Study of Intrinsically Disordered Proteins (IDPs). *ChemBioChem*. 2012; 13(16):2425–2432.

[152] Brutscher B. DEPT spectral editing in HCCONH-type experiments. Application to fast protein backbone and side chain assignment. *Journal of Magnetic Resonance*. 2004;167(2):178–184.

[153] Kupče E, Nishida T, Freeman R. Hadamard NMR spectroscopy. *Progress in Nuclear Magnetic Resonance Spectroscopy*. 2003;42(3-4):95–122.

[154] Brutscher B. Combined frequency- and time-domain NMR spectroscopy. Application to fast protein resonance assignment. *J Biomol NMR*. 2004;29(1):57–64.

[155] Lescop E, Rasia R, Brutscher B. Hadamard amino-acid-type edited NMR experiment for fast protein resonance assignment. *J Am Chem Soc*. 2008;130(15):5014–5015.

[156] Feuerstein S, Plevin MJ, Willbold D, Brutscher B. iHADAMAC: A complementary tool for sequential resonance assignment of globular and highly disordered proteins. *Journal of Magnetic Resonance*. 2012;214(C):329–334.

[157] Pantoja-Uceda D, Santoro J. New amino acid residue type identification experiments valid for protonated and deuterated proteins. *J Biomol NMR*. 2012;54(2):145–153.

[158] Jung YS, Zweckstetter M. Mars - robust automatic backbone assignment of proteins. *J Biomol NMR*. 2004;30(1):11–23.

[159] Lopez-Mendez B, Güntert P. Automated protein structure determination from NMR spectra. *J Am Chem Soc*. 2006;128(40):13112–13122.

[160] Zawadzka-Kazimierczuk A, Koźmiński W, Billeter M. TSAR: a program for automatic resonance assignment using 2D cross-sections of high dimensionality, high-resolution spectra. *J Biomol NMR*. 2012;54(1):81–95.

[161] Kazimierczuk K, Stanek J, Zawadzka-Kazimierczuk A, Koźmiński W. High-dimensional NMR spectra for structural studies of biomolecules. *ChemPhysChem*. 2013;14(13):3015–3025.

[162] Kazimierczuk K, Zawadzka-Kazimierczuk A, Koźmiński W. Non-uniform frequency domain for optimal exploitation of non-uniform sampling. *Journal of Magnetic Resonance*. 2010;205(2):286–292.

[163] Brutscher B, Cordier F, Simorre JP, Caffrey M, Marion D. High-Resolution 3d Hncoca Experiment Applied to a 28-Kda Paramagnetic Protein. *J Biomol NMR*. 1995;5(2):202–206.

[164] Yang DW, Kay LE. TROSY triple-resonance four-dimensional NMR spectroscopy of a 46 ns tumbling protein. *J Am Chem Soc*. 1999;121(11):2571–2575.

[165] Xia YL, Arrowsmith CH, Szyperski T. Novel projected 4D triple resonance experiments for polypeptide backbone chemical shift assignment. *J Biomol NMR*. 2002; 24(1):41–50.

[166] Konrat R, Yang DW, Kay LE. A 4D TROSY-based pulse scheme for correlating 1HNi, 15N, 13Calpha, 13Ci-1 chemical shifts in high molecular weight 15N, 13C, 2H labeled proteins. *J Biomol NMR*. 1999;15(4):309–313.

[167] Staykova DK, Fredriksson J, Bermel W, Billeter M. Assignment of protein NMR spectra based on projections, multi-way decomposition and a fast correlation approach. *J Biomol NMR*. 2008;42(2):87–97.

[168] Clowes RT, Boucher W, Hardman C, Domaille PJ, Laue ED. A 4D HCC(CO)NNH experiment for the correlation of aliphatic side-chain and backbone resonances in 13C/15N-labelled proteins. *J Biomol NMR*. 1993;3(3):349–354.

[169] Mobli M, Stern AS, Bermel W, King GF, Hoch JC. A non-uniformly sampled 4D HCC(CO)NH-TOCSY experiment processed using maximum entropy for rapid protein sidechain assignment. *J Magn Reson*. 2010;204(1):160–164.

[170] Bermel W, Bertini I, Felli IC, Gonnelli L, Koźmiński W, Piai A, Pierattelli R, Stanek J. Speeding up sequence specific assignment of IDPs. *J Biomol NMR*. 2012;53(4):293–301.

[171] Hiller S, Joss R, Wider G. Automated NMR assignment of protein side chain resonances using automated projection spectroscopy (APSY). *J Am Chem Soc*. 2008; 130(36):12073–12079.

[172] Gossert AD, Hiller S, Fernandez C. Automated NMR Resonance Assignment of Large Proteins for Protein-Ligand Interaction Studies. *J Am Chem Soc*. 2011; 133(2):210–213.

[173] Zawadzka-Kazimierczuk A, Kazimierczuk K, Koźmiński W. A set of 4D NMR experiments of enhanced resolution for easy resonance assignment in proteins. *Journal of Magnetic Resonance*. 2010;202(1):109–116.

[174] Nováček J, Haba NY, Chill JH, Žídek L, Sklenář V. 4D non-uniformly sampled HCBCACON and J(NCα)-selective HCBCANCO experiments for the sequential assignment and chemical shift analysis of intrinsically disordered proteins. *J Biomol NMR*. 2012;53(2):139–148.

[175] Shirakawa M, Wälchli M, Shimizu M, Kyogoku Y. The use of heteronuclear cross-polarization for backbone assignment of 2H-, 15N- and 13C-labeled proteins: A pulse scheme for triple-resonance 4D correlation of sequential amide protons and 15N. *J Biomol NMR.* 1995;5(3):323–326.

[176] Piai A, Hošek T, Gonnelli L, Zawadzka-Kazimierczuk A, Koźmiński W, Brutscher B, Bermel W, Pierattelli R, Felli IC. "CON-CON" assignment strategy for highly flexible intrinsically disordered proteins. *J Biomol NMR.* 2014;60(4):209–218.

[177] Zawadzka-Kazimierczuk A, Koźmiński W, Šanderová H, Krásný L. High dimensional and high resolution pulse sequences for backbone resonance assignment of intrinsically disordered proteins. *J Biomol NMR.* 2012;52(4):329–337.

[178] Bermel W, Felli IC, Gonnelli L, Koźmiński W, Piai A, Pierattelli R, Zawadzka-Kazimierczuk A. High-dimensionality C-13 direct-detected NMR experiments for the automatic assignment of intrinsically disordered proteins. *J Biomol NMR.* 2013; 57(4):353–361.

[179] Nováček J, Janda L, Dopitová R, Žídek L, Sklenář V. Efficient protocol for backbone and side-chain assignments of large, intrinsically disordered proteins: transient secondary structure analysis of 49.2 kDa microtubule associated protein 2c. *J Biomol NMR.* 2013;56(4):291–301.

[180] Mukrasch MD, Bibow S, Korukottu J, Jeganathan S, Biernat J, Griesinger C, Mandelkow E, Zweckstetter M. Structural polymorphism of 441-residue tau at single residue resolution. *PLoS Biol.* 2009;7(2):e34.

[181] Hiller S, Wasmer C, Wider G, Wüthrich K. Sequence-specific resonance assignment of soluble nonglobular proteins by 7D APSY-NMR spectroscopy. *J Am Chem Soc.* 2007;129(35):10823–10828.

[182] Kay LE, Ikura M, Zhu G, Bax A. Four-dimensional heteronuclear triple-resonance NMR of isotopically enriched proteins for sequential assignment of backbone atoms. *J Magn Reson.* 1991;91(2):422–428.

[183] Boucher W, Laue ED, Campbell-Burk SL, Domaille PJ. Improved 4D NMR experiments for the assignment of backbone nuclei in13C/15N labelled proteins. *J Biomol NMR.* 1992;2(6):631–637.

[184] Szyperski T, Wider G, Bushweller JH, Wüthrich K. 3D 13C-15N-heteronuclear two-spin coherence spectroscopy for polypeptide backbone assignments in 13C-15N-double-labeled proteins. *J Biomol NMR*. 1993;3(1):127–132.

[185] Löhr F, Rüterjans H. A new triple-resonance experiment for the sequential assignment of backbone resonances in proteins. *J Biomol NMR*. 1995;6(2):189–197.

[186] Bagai I, Ragsdale SW, Zuiderweg ERP. Pseudo-4D triple resonance experiments to resolve HN overlap in the backbone assignment of unfolded proteins. *J Biomol NMR*. 2011;49(2):69–74.

[187] Hiller S, Fiorito F, Wüthrich K, Wider G. Automated projection spectroscopy (APSY). *Proc Natl Acad Sci USA*. 2005;102(31):10876–10881.

[188] Malmodin D, Billeter M. Multiway decomposition of NMR spectra with coupled evolution periods. *J Am Chem Soc*. 2005;127(39):13486–13487.

[189] Kim S, Szyperski T. GFT NMR, a new approach to rapidly obtain precise high-dimensional NMR spectral information. *J Am Chem Soc*. 2003;125(5):1385–1393.

[190] Kim S, Szyperski T. GFT NMR experiments for polypeptide backbone and 13Cbeta chemical shift assignment. *J Biomol NMR*. 2004;28(2):117–130.

[191] Arnesano F, Banci L, Bertini I, Cantini F, Ciofi-Baffoni S, Huffman DL, O'Halloran TV. Characterization of the binding interface between the copper chaperone Atx1 and the first cytosolic domain of Ccc2 ATPase. *J Biol Chem*. 2001;276(44):41365–41376.

[192] Arnesano F, Balatri E, Banci L, Bertini I, Winge DR. Folding studies of Cox17 reveal an important interplay of cysteine oxidation and copper binding. *Structure/Folding and Design*. 2005;13(5):713–722.

[193] Bertini I, Felli IC, Gonnelli L, Kumar V M V, Pierattelli R. 13C Direct-Detection Biomolecular NMR Spectroscopy in Living Cells. *Angew Chem Int Ed*. 2011; 50(10):2339–2341.

[194] Csizmok V, Felli IC, Tompa P, Banci L, Bertini I. Structural and Dynamic Characterization of Intrinsically Disordered Human Securin by NMR Spectroscopy. *J Am Chem Soc*. 2008;130(50):16873–16879.

5.2 Application of NMR methods to viral IDPs

5.2.1

The heterogeneous structural behavior of E7 from HPV 16 revealed by NMR spectroscopy

Eduardo O. Calçada[a], Isabella C. Felli[a], Tomáš Hošek[a], Roberta Pierattelli[a]

[a] *CERM and Department of Chemistry Ugo Schiff, University of Florence, via Luigi Sacconi 6, Sesto Fiorentino, 50019, Italy*

Published in **2013** in *ChemBioChem*, 14(14), 1876-82

ChemPubSoc Europe

DOI: 10.1002/cbic.201300172

The Heterogeneous Structural Behavior of E7 from HPV16 Revealed by NMR Spectroscopy

Eduardo O. Calçada,[a] Isabella C. Felli,*[a, b] Tomáš Hošek,[a] and Roberta Pierattelli*[a, b]

This paper is dedicated to the memory of our Maestro, Prof. Ivano Bertini

The E7 protein from human papillomavirus (HPV) plays a key role in oncogenesis; for this reason, it is a target of great biomedical interest. To date, no high resolution information is available for the full protein. We present here the NMR characterization of the entire E7 from HPV16, one of the most oncogenic variants of the virus. The protein is very heterogeneous in terms of structural and dynamic properties with a highly flexible N-terminal module and a more structured C terminus. This opens possibilities for studies of molecular-level interactions and post-translational modifications of the protein to unravel functional details that might be linked to its highly oncogenic potential.

Introduction

The importance of local flexibility in determining the function of proteins was recognized long ago and has been widely scrutinized.[1–5] If the extent of local flexibility is taken to extremes, it leads to completely random coil behavior of polypeptide chains, indicated as intrinsic disorder,[6–8] through a wide variety of intermediate cases, both in terms of extent of mobility and in terms of protein stretches involved. Intrinsically disordered proteins (IDPs) possess complementary properties to well-folded ones. Many examples of IDPs have appeared in the literature, showing how their structural plasticity, adaptability, and intrinsic flexibility can actually be key features that enable them to interact with a variety of different partners, adapt to different conditions, and act as hub proteins.[9,10] These peculiar properties provide functional advantages to IDPs, enabling them to play key roles in many regulatory processes.[11] Their function/malfunction has also been related to several diseases.[12]

Thanks to the largely extended and exposed backbones, IDPs often display short amino acid segments (short linear motives, SLiMs) that might interact with different partners.[13,14] This strategy of exploiting proteins' SLiMs for driving interactions, seems to be extensively used by viruses to infect healthy cells. In virtue of their small genomes, able to code only a limited number of proteins, they need economic ways to interfere with the host.[13] This is the case of human papilloma virus (HPV).[15]

HPV is a family of more than 100 different viruses.[16] They can be clinically classified as "low risk" (LR), causing benign warts, and as "high risk" (HR), which are causative agents of cervical cancer, the second most common cancer in women worldwide.[17] The entire proteome of HPV includes six non-structural proteins (E1, E2, E4, E5, E6, and E7) and two structural proteins (L1 and L2). The oncogenic potential of HR HPV relies on the expression of E6 and E7, which are consistently present in HPV-associated carcinomas and are necessary for induction and maintenance of the transformed phenotype.[18–20]

E7 is a small protein comprising about 100 amino acids. Three conserved regions (CR), namely CR1 and CR2 in the N-terminal half and CR3 in the C-terminal half,[21–23] have been identified (Figure 1). The first two (CR1 and CR2), which have been found to interact with a high variety of target proteins,[19] show a high propensity to be disordered.[24] The CR3 region, which contains two CXXC motifs separated by 29 to 30 amino acids, has been shown to bind zinc and is expected to be more structured.[24]

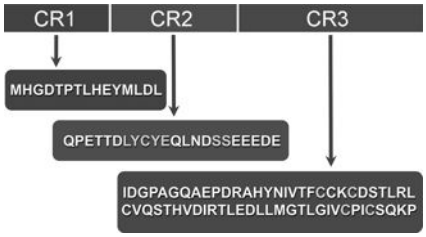


Figure 1. Schematic representation of HPV16 E7 oncoprotein amino acid sequence, and division into three conserved regions.

Structural characterization of the CR3 of the protein, which contains zinc binding motifs, was achieved for short constructs comprising only CR3 for the HPV1a variant through X-ray[25]

[a] E. O. Calçada,[+] Prof. I. C. Felli, T. Hošek,[+] Prof. R. Pierattelli
Magnetic Resonance Center (CERM), University of Florence
Via Luigi Sacconi 6, 50019 Sesto Fiorentino (Italy)
E-mail: felli@cerm.unifi.it
pierattelli@cerm.unifi.it

[b] Prof. I. C. Felli, Prof. R. Pierattelli
Department of Chemistry "Ugo Schiff", University of Florence
Via della Lastruccia 3, 50019 Sesto Fiorentino (Italy)

[+] These authors contributed equally to this work.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/cbic.201300172.

and for the HPV45 variant through nuclear magnetic resonance (NMR) spectroscopy.[26] However, the whole protein failed to give crystals and/or was not characterized through NMR. Atomic resolution information for HPV16 E7 is still lacking,[26] despite the interest in accessing high resolution information for the protein responsible for the high oncogenic potential of this variant and its interactions.

Among the many documented interactions of E7 with host proteins,[15] the one with the retinoblastoma tumor suppressor (pRB) plays a key role in the oncogenic potential of E7 and has been extensively characterized by many techniques.[27] The structure of a short polypeptide containing the LXCXE motif localized in the CR2 region[28] and bound to pRB is available, providing precious hints toward understanding how this short fragment binds to pRB.[29] However, this polypeptide has much lower affinity for pRB with respect to the whole protein,[28] indicating that other parts of the protein also contribute to binding.[30] E7 binds to several other partners, increasing their activity to enhance pRB degradation. The cullin 2 ubiquitin ligase complex is bound by the CR1 of the E7 protein.[31] The CR3 also interacts with the subunit 4 (S4) ATPase of the S26 proteasome and enhances its APTase activity, providing another pathway for pRB degradation.[32] E7 also interacts through its N terminus with other pRB-related tumor-suppressor proteins, p107 and p130.[33,34] pRB degradation is just one example of the complex array of interactions established by viruses like HPV to hijack cell regulation and ensure survival.

To contribute to the understanding of the sophisticated mechanisms behind HPV-related diseases, it is of primary importance to access high resolution structural and dynamic information of the small, though complex, oncogenic proteins of HPV. In this work, we present the characterization by NMR of the full construct of E7 from HPV16. Despite a small molecular mass, its heterogeneous nature limited the possibility of accomplishing the complete sequence-specific assignment but provided clear hints regarding its structural and dynamic features.

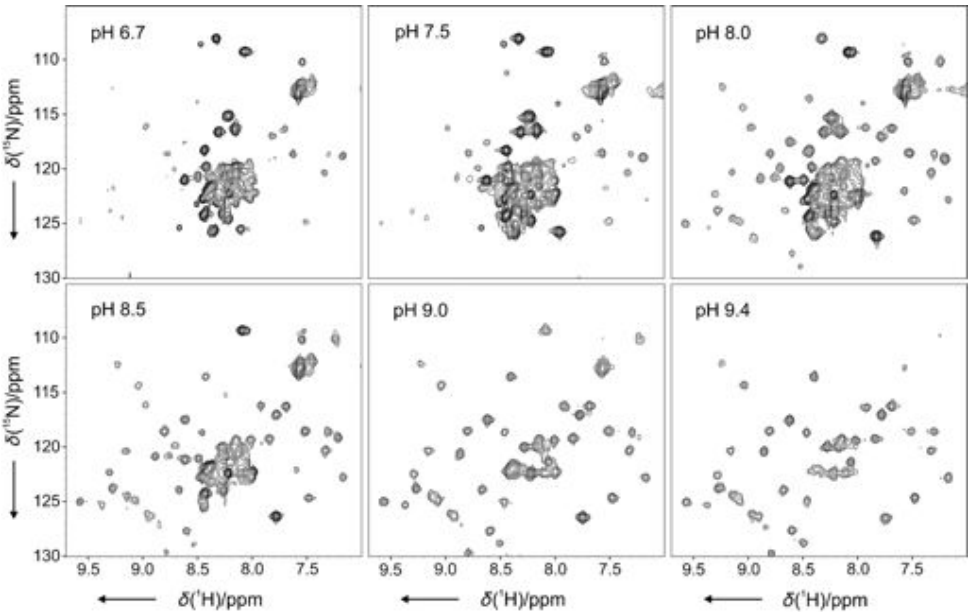


Figure 2. Portions of the ^{1}H–^{15}N SOFAST-HMQC spectra recorded for HPV16 E7 at pH 6.7, 7.0, 7.5, 8.0, 8.5, and 9.4. Spectra have been processed to maximize signal-to-noise at the expense of resolution. The higher signal-to-noise ratio for the set of strong signals clustered in the central region was obtained at pH 7.5. The second set of signals, characterized by a larger chemical shift dispersion, is most intense at pH 8.5. Spectra were recorded at 18.8 T and 298 K for a 0.3 mM sample in 10 mM HEPES buffer, 150 mM KCl, 10 μM $ZnCl_2$, and 10 mM DTT.

Results

The heterogeneous structural and dynamic properties of HPV16 E7 are evident from the 2D ^{1}H–^{15}N correlation NMR spectra reported in Figure 2. The figure shows two distinct sets of signals with very different properties in terms of chemical shift dispersion, signal-to-noise ratio (S/N), and line width. In the first set, clustered in a very narrow spectral region (7.9–8.6 ppm in the ^{1}H dimension), signals are characterized by relatively sharp lines and high S/N and tend to disappear with increasing pH. These are all distinctive properties of signals belonging to nuclei of a highly flexible, disordered, and solvent-exposed module. Elevated exchange rates of amide protons with solvent, typical of unfolded polypeptides, justify the disappearance of the signal with increasing pH, broadening lines beyond detection as observed in the present case.

A second set of much weaker and broader signals, characterized by a larger chemical shift dispersion (7.0–9.6 ppm in the ^{1}H dimension), becomes detectable with increasing pH (Figure 2, pH 8.0–9.4). This second set of signals becomes dominant with increasing pH, whereas the first set decreases in intensity and eventually becomes undetectable (Figure 2, pH > 9.0). The large chemical shift dispersion of this second set of signals is a clear indication that they belong to residues of a more structured part of the protein, also consistent with amide protons in a more solvent-protected conformation, which remain detectable up to pH 9.8. However, line widths of the signals are much larger with respect to what is expected for a well-structured, well-behaved, globular protein of small size. Therefore, we ascribe this second set of signals to a more structured part of the protein which does experience extensive line broadening. The extent of disorder of different parts of the protein from different variants of the virus, as predicted through bioinformatic tools, has been linked to their oncogenic potential.[24]

Sequence-specific assignment of the first set of signals is shown in Figure 3. This was obtained through a combination of ^{1}H- and ^{13}C-detected multidimensional NMR experiments, which enabled us to unambiguously assign 95% of the resi-

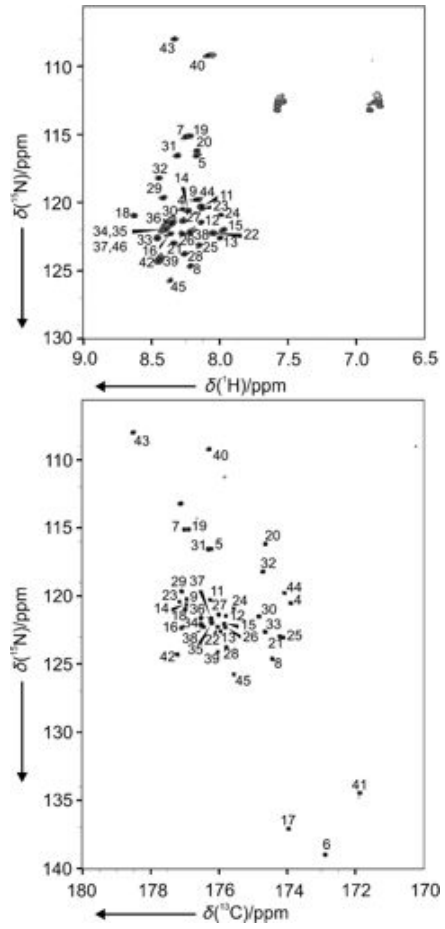


Figure 3. $^{1}H–^{15}N$ SOFAST-HMQC and $^{13}C–^{15}N$ CON-IPAP spectra recorded for HPV16 E7. The resulting assignment is reported in the figure, indicating the residue number (relative to the backbone nitrogen). The complete chemical shifts table is reported in Table S1. Experiments were recorded at 16.4 T and 298 K for a 0.76 mM E7 sample in 10 mM HEPES buffer at pH 7.5, 150 mM KCl, 10 μM $ZnCl_2$, and 10 mM DTT.

dues in the N terminus of the protein (E7N, residues 1–46) comprising CR1 and CR2. Chemical shifts of H^N, N, C′, C^α, and C^β of the assigned residues are reported in Table S1 of the Supporting Information and were deposited in the BMRB. The assignment, originally achieved at pH 7.5, was repeated at pH 8.5 to evaluate possible chemical shift changes of the already assigned crosspeaks and to try to detect a sufficient number of correlations involving the second set of crosspeaks to enable their sequence-specific assignment. However, whereas on one hand it was possible to easily transfer the assignments of the crosspeaks from pH 7.5 to pH 8.5, the sensitivity of the triple resonance experiments was not sufficient to perform the sequence-specific assignment of the second set of crosspeaks identified in 2D $^{1}H–^{15}N$ maps, even though several variants were tested, including those that were relaxation optimized.

The second set of signals could also, in principle, be due to a second conformation of E7N. To exclude this possibility and confirm that the second set of observed signals derives from the final part of the protein comprising CR3, a shorter E7 construct (residues 45–98, E7C) was expressed and purified. The recorded 2D $^{1}H–^{15}N$ correlation map of E7C (Figure S1) confirmed that the second set of signals observed in analogous spectra recorded on the full construct was due to the final module of the protein (45–98, E7C). This comparison also shows that general spectral characteristics of E7C alone and of E7C in the full construct are very similar.

Chemical shifts are the first reporters of the structural and dynamic properties of a protein that can be analyzed after sequence-specific assignment. Indeed, analysis of experimental heteronuclear chemical shifts (N, C′, C^α, and C^β) and comparison with the so-called random-coil chemical shift values can be used to investigate local secondary structural propensities and ascertain to what extent a certain local conformation is populated. The secondary structure propensity (SSP) score,[35] which combines the different chemical shifts into a single residue-specific parameter, is shown in Figure 4.

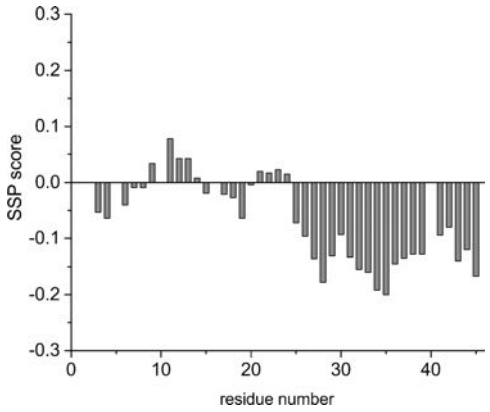


Figure 4. Secondary structure propensity (SSP) score,[31] obtained by combining N^H, C′, C^α, and C^β chemical shifts for the assigned resonances of HPV16 E7.

To have a more general picture of the structural and dynamic properties of the protein through NMR, additional observables were determined. The most relevant were the ^{15}N R_1, ^{15}N R_2, and $^{1}H–^{15}N$ NOEs, which can be used to identify parts of the polypeptide chain characterized by different motional properties. The magnitude of the heteronuclear $^{1}H–^{15}N$ NOE effects is one of the most informative measures, as it depends on the local effective correlation time and is very sensitive to fast local motions. Transverse relaxation rates are also sensitive to exchange processes. The combined analysis of these data can be used for characterization of the protein. These data, measured at pH 7.5, are reported in Figure 5 both for the assigned E7N regions (left panels) as well as for the unassigned E7C regions (right panels). The latter are ordered according to increasing ^{15}N R_2 values to facilitate comparison.

All the NMR data determined for the E7N region indicate that it is highly flexible and highly disordered, as is also immediately evident from the SSP values as well as from negative $^{1}H–^{15}N$ heteronuclear NOE. Detailed inspection of these data also indicates short fragments of amino acids characterized by a mild secondary structural propensity, starting at residue 26 onwards (Figure 4). From a general point of view, the deter-

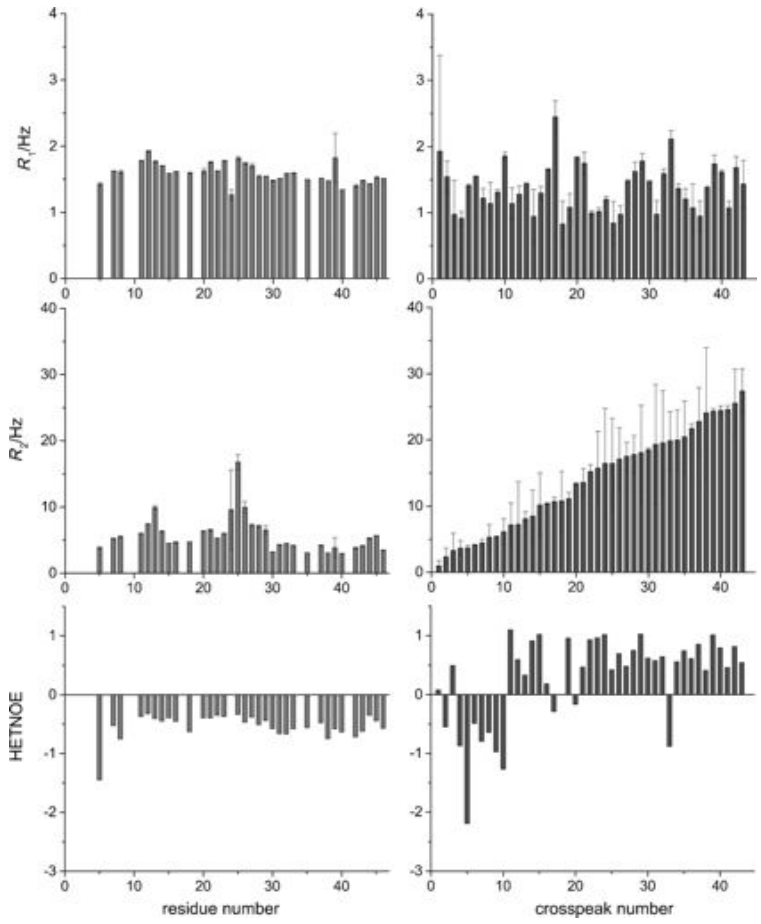


Figure 5. Relaxation data for backbone amide ^{15}N nuclei of HPV16 E7 as a function of the residue number for the N-terminal region (E7N) and of the peak number for the C-terminal region (E7C), with the latter ordered according to increasing R_2 values. From top to bottom: ^{15}N longitudinal relaxation rates (R_1), ^{15}N transverse relaxation rates (R_2), and ^1H–^{15}N NOE values (HETNOE). Experiments were recorded at 11.8 T and 298 K for a 0.25 mM sample in HEPES buffer at pH 7.5, 150 mM KCl, 10 μM $ZnCl_2$, and 10 mM DTT.

mined data agree with the overall picture of a protein characterized by a very flexible N terminus and a more structured C terminus which, however, is characterized by transverse relaxation rates that are much larger than expected for a protein of about 50 amino acids. Calculation of ^{15}N relaxation rates through the HYDRONMR program,[36] by using the monomeric (PDB ID: 2EWL) and dimeric (PDB ID: 2F8B) CR3 NMR structures of the HPV45 E7[26] variant as input (see the Supporting Information), indicates the presence of dimeric forms of CR3. Furthermore, some residues in CR3 display even higher R_2 value than expected for a dimeric form. This could be due to conformational exchange processes deriving from a "loose", molten globule-like structure, or from the tendency of the protein to aggregate and form high molecular mass species in solution in equilibrium with monomeric or dimeric forms. All structural evidence available on the constructs of the C terminus from different viral types shows that the CR3 region tends to form dimers,[25,26] and biophysical data support the tendency of the CR3 to form oligomers.[37,38] The latter tendency might be also responsible for the disappearance of the CR3 peaks at low pH.[37,38] In order to gain some insights into possible origins of the observed broadening involving the C terminus of the protein, concentration-dependent measurements and temperature-dependent measurements were collected.

Concentration dependence was monitored through a series of 2D ^1H–^{15}N correlation maps at pH 7.5 (Figure 6). This figure reports the ratios of intensities of selected crosspeaks, accurately chosen to be representative of the E7N as well as of the more structured E7C regions of the protein. From the average intensity ratio, a function of the protein concentration for the two regions of the protein (right panel), it is clearly evident that the intensity of signals in E7N increases linearly with the concentration, whereas for the other region, this ratio increases much more slowly with the concentration in agreement with the hypothesis of the formation of high molecular mass species that involve the E7C, leaving the E7N highly flexible and thus still detectable by NMR.

The temperature dependence of the selected crosspeaks shows that the signals deriving from the E7C irreversibly disappear when the temperature is increased above 35 °C, in agreement with the formation of high-order aggregates that are not detectable by NMR spectroscopy. Indeed, upon increase in temperature and loss of the E7C signals, no precipitate is observed at either

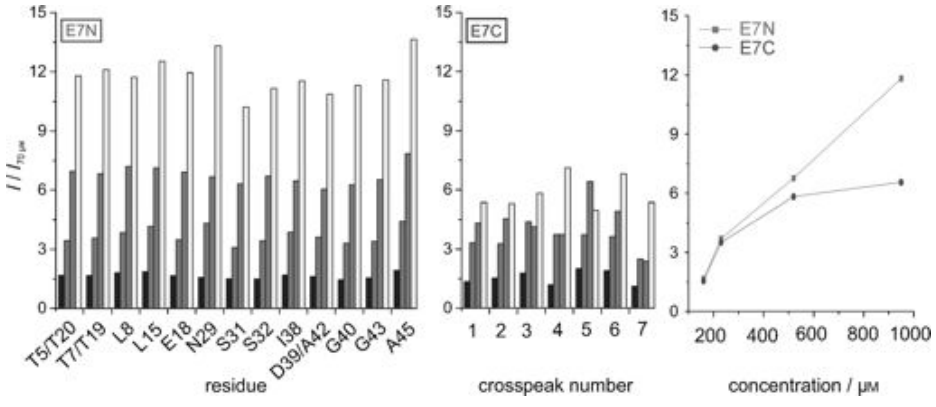


Figure 6. Diagram of the intensity of selected ^1H–^{15}N SOFAST-HMQC signals for HPV16 E7 recorded at various concentrations of protein: 160, 230, 520, and 950 μM. Data are scaled to the intensity of the signals from a 70 μM sample. The two sets of signals were selected to be representative of the E7N (left) and of the E7C (right) regions of the protein. In the right panel, the average intensity ratio as a function of the protein concentration for the two parts of the protein are compared. The experiments were recorded at 18.8 T and 298 K for samples in 10 mM HEPES buffer at pH 7.5, 50 mM KCl, 10 μM $ZnCl_2$, and 10 mM DTT.

pH 7.5 or 8.5, likely due to the fact that E7N at this pH is highly charged and responsible for the high solubility of this protein. The set of signals from E7N is still detectable at temperatures above 35 °C, even at high temperature as observed for other disordered proteins.[39]

Many different experimental conditions were tested through 2D $^{1}H–^{15}N$ NMR spectra in order to evaluate possible changes in the observed crosspeaks upon changes in ionic strength, buffer, and type and concentration of reducing agent, as well as to improve the S/N and overall relaxation properties of E7C signals. However, the overall behavior of the protein was similar and in line with the presence of a highly flexible N-terminal region, indicating the stability of this protein element under a variety of different conditions. On the other hand, the signals deriving from the C-terminal region of the protein were detectable under specific conditions and were very sensitive to increasing temperature but also resistant to denaturation (not shown).

Discussion

The NMR data obtained for HPV16 E7 are in line with a protein characterized by very heterogeneous structural and dynamic behavior, with a highly flexible N-terminal module (E7N) and a more structured C-terminal region (E7C). The N terminus comprises the so-called CR1 and CR2, as identified by comparison of the primary sequences of E7 from different types of viruses. This part of the protein does contain a few short linear motifs, which are involved in interactions with protein partners. These include the LXCXE motif (22–26), identified as the key motif for interaction with pRB,[28,29] and the SS motif (31–32), the known site of phosphorylation by casein kinase II.[40] Atomic resolution information for this part of the protein is not available for any virus variant; thus, the current data can provide detailed information useful for characterizing the structural and dynamic properties of this part of the polypeptide sequence that, until now, has been broadly defined as disordered. Inspection of NMR data confirms that the N-terminal region is characterized by high flexibility, as evident from the negative values of $^{1}H–^{15}N$ NOEs and chemical shift values near those of random-coil. The $^{1}H–^{15}N$ NOEs do approach zero, progressing with the amino acid sequence, and eventually become positive as they approach the C-terminal region of the protein, indicating a less flexible behavior.

The part of the polypeptide sequence from residues 26–46 results from NMR data characterized by a mild propensity (~10%) to adopt an elongated structured conformation. This is the fragment immediately following the LXCXE motif and containing the SS phosphorylation site. It is interesting to note that the LXCXE motif in a nine-residue peptide mimicking the E7 interacting sequence has been shown to bind pRb in a β-strand-like conformation.[28] The propensity to adopt an elongated conformation, identified here for the residues immediately following this fragment, could be relevant to promote binding and also include the phosphorylation site.

The highly charged E7N, responsible for the high solubility of the protein near neutral pH, could also be responsible for preventing the interaction of the N-terminal with the C-terminal regions of E7, maintaining E7N in a fairly elongated form to promote interaction with partners and to ensure that the kinase motif is available for post-translational modifications. Moreover, keeping the two modules apart might be an important feature to render them "independent" one from the other but still synergistic, in order to provide two independent units to modulate the interaction with other partners but that remain next to each other. The availability of the sequence-specific assignment of E7N opens the path to atomic resolution characterization of its interaction with partners.

On the other hand, as stated above, no assignment could be obtained for the C-terminal part of the protein. However, a large set of crosspeaks from this part of the protein was detected in $^{1}H–^{15}N$ correlation experiments and can be used to determine overall structural and dynamic properties of this module of the protein. The large chemical shift dispersion of the signals is a clear indicator of a more structured polypeptide which, however, is very dissimilar to a folded protein of similar size, due to its markedly different relaxation properties. An increase in transverse relaxation of this module, monitored both from ^{1}H and ^{15}N line widths as well as by ^{15}N transverse relaxation rates, could arise from either conformational exchange broadening, due to a loosely defined tertiary or quaternary structure which affects nuclear relaxation rates but is not loose enough to affect chemical shifts, or by formation of higher molecular mass species in equilibrium in solution, or by an interaction with the N-terminal flexible module. ^{15}N relaxation rates, in particular for $^{1}H–^{15}N$ NOEs, indicate that the structure is not fixed but that it is characterized by a significant degree of local motion, in agreement with the presence of large local fluctuations that would indicate an initial molten globule state. The concentration dependence of the different sets of signals also indicates aggregation of the C-terminal region while retaining a very flexible N-terminal region. This would provide an interesting mechanism to increase the effective concentration of the N-terminal region by using the C-terminal region to group different molecules together, leaving the N terminus free to move and thus interact with potential partners. The only hypothesis that does not seem to be relevant is that broadening of the C-terminal region is due to mutual crosstalk between the N terminus and C terminus, as the N-terminal region should also be affected in this case. Instead, as mentioned above, the N-terminal region of the protein under conditions close to physiological conditions, seems to provide the driving force for the high solubility of the protein and does remain structurally disordered under a variety of different experimental conditions.

Detection of the signals of the C-terminal region of the protein, even in the absence of their sequence-specific assignment, might also enable us to detect possible structural changes involving this part of the protein upon interaction.[23,41] Indeed, it has been shown that it is important to evaluate the interactions of E7 with other proteins in its full form, as different components appear to have complementary roles, and shorter constructs display lower affinities.[18,29,42] Further work is in progress through different NMR-based strat-

egies to extend assignment to the C-terminal region of the protein.

Conclusions

The thorough NMR investigation of E7 from HPV16 presented here reveals very heterogeneous structural and dynamic properties, presenting a variable extent of disorder which could be important for the function of the protein, well beyond the accepted schematic view of CR1 and CR2 being fully disordered and CR3 well-structured. In addition, this NMR study provides an important platform for the study of the post-translational modifications and interactions of E7 from HPV16 with the many partners known to interact with different parts of the protein. In conclusion, E7 from HPV16, constituted of <100 amino acids, is able to efficiently interact with several key protein partners, interfering with crucial cell-regulatory processes. The large structural and dynamic heterogeneity monitored here by NMR spectroscopy might be an important aspect in this respect, and the current study opens the possibility for atomic resolution characterization of these interactions, which could provide important information to understand the features that promote progression of cells to malignancy[20,43] and possibly to design new drugs.[44]

Experimental Section

Construction of bacterial expression vectors: The recombinant vector pET20-E7, containing the E7 gene from HPV16, was generously provided by Scil Proteins GmbH.[45] As previously described,[45] the gene was modified to generate silent mutations in codons for His2, to eliminate the possibility of alternative translation initiation and premature termination of translation, and for Pro6, to increase the protein expression yield (pET20-E7-opt). The optimized plasmid construction was transformed into *E. coli* XL1-Blue competent cells (Stratagene), and the resulting clones were screened by DNA sequence analysis of miniprep-purified plasmid samples.

The DNA encoding HPV16 E7 CR3 (45–98 aa) was amplified by PCR from the pET20-E7 plasmid encoding the full-length protein and inserted into the Gateway pEntr-TEV-d-Topo vector to create an entry clone. The entry clone containing the gene of interest was transferred into N-terminal fusion pDEST17 vector containing a His_6 tag by a Gateway LR Clonase II recombination reaction to create the expression clone pDEST17-E7CR3.

Proteins expression and purification: The recombinant vector pET20-E7-opt was transformed into *E. coli* BL21(DE3)pLysS (Stratagene) for protein expression. One colony was selected to inoculate LB medium (50 mL) containing ampicillin and chloramphenicol grown overnight at 37 °C and 180 rpm. For ^{15}N- and ^{13}C,^{15}N-labeled protein production, a preinoculum (10 mL) was used to inoculate minimal medium (1 L, 48.5 mM Na_2HPO_4, 22.0 mM K_2HPO_4, 8.5 mM NaCl, 0.2 mM $CaCl_2$, 2.0 mM $MgSO_4$, 10.0 μM $ZnCl_2$, 1 mg L^{-1} each of biotin and thiamin, 7.5 mM $(^{15}NH_4)_2SO_4$, and 11.1 mM glucose/$^{13}C_6$-glucose) at 37 °C with constant agitation at 160 rpm. When the OD_{600} reached 0.6, cells were induced with IPTG (1 mM final concentration) and allowed to grow for an additional 4 h at 30 °C. The cells were harvested by centrifugation at 8000 *g* for 20 min, and the cell pellet was stored at −20 °C.

Frozen cells were thawed and suspended in 40 mL of equilibration buffer A (50 mM NaH_2PO_4, 300 mM NaCl, and 10 mM imidazole, pH 8.0). Cells were disrupted by sonication on ice (at 40% sonication power) with cycles of 3 s with 10 s delay pulses for 10 min.

Lysed cells were centrifuged at 165 000 *g* for 30 min at 4 °C, and the supernatant was loaded on a 5 mL Ni^{2+} HiTrap column (Amersham Biosciences), pre-equilibrated with buffer A under anaerobic conditions. The column was washed with 10 column volumes of buffer A containing imidazole (50 mM). The E7 protein was eluted with buffer A (40 mL) containing imidazole (250 mM) and then dialyzed against HEPES (10 mM) and KCl (50 mM, pH 7.5). Samples were concentrated to 3.0 mL in the presence of dithiothreitol (DTT; 10 mM) and were stored at 4 °C for biophysical analyses. Mass-spectrometry (MALDI and ESI) was performed to confirm the molecular mass of purified E7 protein. All samples for NMR experiments were prepared in HEPES buffer (10 mM) containing KCl (150 mM), $ZnCl_2$ (10 μM), and DTT (10 mM).

For the CR3 construct, the vector was expressed in BL21 (DE3)-pLysS cells, and expression was conducted as for the full-length construct. After harvesting and lysing the cells, the protein remained in the insoluble phase as inclusion bodies. The pellet was suspended in buffer B (8 M urea, 5 mM imidazole, 5 mM DTT, 10 μM $ZnCl_2$, and 10 mM HEPES, pH 8.5) and loaded on Ni^{2+} HiTrap column pre-equilibrated with same buffer. The column was washed with 10 column volumes of buffer B containing imidazole (10 mM). The E7 CR3 fused protein was eluted with buffer B containing imidazole (50 mM) and concentrated to 500 μL and a concentration of 30 mg mL^{-1} for further refolding. Refolding was performed inside a glove box by drip dilution at one drop every 30 s into 50 mL degassed buffer C (10 mM HEPES, 50 mM KCl, 10 mM DTT, and 10 μM $ZnCl_2$, pH 8.5) to achieve 100-fold dilution and a final protein concentration of 0.3 mg mL^{-1}. Refolded protein was concentrated to 1 mL and buffer-exchanged into TEV buffer (20 mM Tris pH 8.0, 50 mM KCl, 5 mM citrate, 10 μM $ZnCl_2$, and 0.5 mM TCEP). The His tag was cleaved from the recombinant protein by incubation with TEV protease and was removed by reverse Ni^{2+} affinity chromatography in HEPES (10 mM, pH 7.5) containing KCl (50 mM) and $ZnCl_2$ (10 μM). Cleaved, pure E7 CR3 protein was concentrated to 300 μL with a final concentration of 40 μM.

NMR experiments: For determination of optimal sample conditions, different variants of 2D ^1H–^{15}N experiments (SOFAST-MQC,[46] HSQC[47,48]) were carried out on Bruker DRX 500 and Bruker Avance 700, 800, and 900 instruments, all equipped with cryogenically cooled probe heads.

The ^1H-detected NMR experiments for sequence-specific assignment were acquired at 298 K on the 18.8 T Bruker Avance 800, operating at 800.13 MHz for ^1H, equipped with a cryogenically cooled probe. ^{13}C-detected, exclusively heteronuclear, NMR experiments for sequence-specific assignment were carried out at 298 K on a 16.4 T Bruker Avance 700, operating at 176.03 MHz for ^{13}C, equipped with cryogenically cooled probe optimized for ^{13}C sensitivity. A dataset of ^1H-detected (SOFAST HMQC,[46] HNCO,[49,50] HNCACO,[51] CBCACONH,[52] HNCACB,[53] and HN(CA)NNH[54]) and ^{13}C-detected (protonless CON-IPAP,[55] HCACO-IPAP,[56,57] and HCBCACON-IPAP[57]) NMR experiments was necessary for sequence-specific assignment of H^N, N, C′, C^α, and C^β resonances.

To further characterize the structural and dynamic features of HPV16 E7 oncoprotein, the following ^1H-detected NMR experiments were acquired at 298 K: ^{15}N relaxation experiments (R_1, R_2, and ^1H–^{15}N NOE)[5,58] experiments on an 11.7 T Bruker DRX 500

spectrometer equipped with a cryogenically cooled triple-resonance probe head. Experimental details are reported in Table S2.

Data were processed with TopSpin 2.0 and were analyzed with the programs CARA[59] and Sparky.[60] The secondary structure propensity from chemical shifts was determined by using the SSP program.[35] The ^{15}N relaxation rates (R_1 and R_2) were determined by fitting the cross peak volumes, measured as a function of variable delay, to single-exponential decay. ^{1}H–^{15}N NOE values were obtained as a ratio between peak volumes in spectra recorded with and without ^{1}H saturation. Relaxation rate calculation was performed by using HYDRONMR[36] version 7C.

Acknowledgements

This work was supported in part by the Joint Research Activity and Access to Research Infrastructures (BioNMR, contract 261863) and by the Marie Curie ITN programs (IDPbyNMR, contract 264257) in the EC 7th Framework.

Keywords: heteronuclear NMR · HPV · IDP · viral proteins

[1] G. Wagner, S. Hyberts, J. W. Peng in *NMR of Proteins* (Eds.: G. M. Glore, A. M. Gronenborn), CRC Press, New York, **1993**, pp. 220 – 257.
[2] G. Wagner, *Curr. Opin. Struct. Biol.* **1993**, *3*, 748 – 754.
[3] A. Mittermaier, L. E. Kay, *Science* **2006**, *312*, 224 – 228.
[4] A. Mittermaier, L. E. Kay, *Trends Biochem. Sci.* **2009**, *34*, 601 – 611.
[5] G. Barbato, M. Ikura, L. E. Kay, R. W. Pastor, A. Bax, *Biochemistry* **1992**, *31*, 5269 – 5278.
[6] I. C. Felli, R. Pierattelli, P. Tompa in *NMR of Biomolecules: Towards Mechanistic Systems Biology* (Eds.: I. Bertini, K. S. McGreevy, G. Parigi), Wiley-Blackwell, Weinheim, **2012**, pp. 137 – 152.
[7] P. Tompa, *Trends Biochem. Sci.* **2012**, *37*, 509 – 516.
[8] B. He, K. Wang, Y. Liu, B. Xue, V. Uversky, A. K. Dunker, *Cell Res.* **2009**, *19*, 929 – 949.
[9] P. E. Wright, H. J. Dyson, *J. Mol. Biol.* **1999**, *293*, 321 – 331.
[10] V. N. Uversky, C. J. Oldfield, A. K. Dunker, *J. Mol. Recognit.* **2005**, *18*, 343 – 384.
[11] M. Fuxreiter, I. Simon, P. Friedrich, P. Tompa, *J. Mol. Biol.* **2004**, *338*, 1015 – 1026.
[12] V. Uversky, C. J. Oldfield, A. K. Dunker, *Annu. Rev. Biophys.* **2008**, *37*, 215 – 246.
[13] N. E. Davey, G. Trave, T. J. Gibson, *Trends Biochem. Sci.* **2011**, *36*, 159 – 169.
[14] P. Tompa, M. Fuxreiter, *Trends Biochem. Sci.* **2008**, *33*, 2 – 8.
[15] L. B. Chemes, J. Glavina, J. Faivovich, G. de Prat-Gay, I. E. Sánchez, *J. Mol. Biol.* **2012**, *422*, 336 – 346.
[16] H. U. Bernard, R. D. Burk, K. van Doorslaer, H. zur Hausen, E. M. de Villiers, *Virology* **2010**, *401*, 70 – 79.
[17] H. zur Hausen, *Nat. Rev. Cancer* **2002**, *2*, 342 – 350.
[18] M. G. Donà in *Oncogene Proteins: New Research* (Eds.: A. H. Malloy, E. C. Carson), Nova Science, Hauppauge, **2008**, pp. 19 – 63.
[19] M. E. McLaughlin-Drubin, K. Münger, *Virology* **2009**, *384*, 335 – 344.
[20] L. Banks, D. Pim, M. Thomas, *Nat. Rev. Cancer* **2012**, *12*, 877 – 886.
[21] J. R. Gage, C. Meyers, F. O. Wettstein, *J. Virol.* **1990**, *64*, 723 – 730.
[22] W. C. Phelps, K. Münger, C. L. Yee, J. A. Barnes, P. M. Howley, *J. Virol.* **1992**, *66*, 2418 – 2427.
[23] B. Todorovic, K. Hung, P. Massimi, N. Avvakumov, F. A. Dick, G. S. Shaw, L. Banks, J. S. Mymryk, *J. Virol.* **2012**, *86*, 13313 – 13323.
[24] V. N. Uversky, A. Roman, C. J. Oldfield, A. K. Dunker, *J. Proteome Res.* **2006**, *5*, 1829 – 1842.
[25] X. Liu, A. Clements, K. Zhao, R. Marmorstein, *J. Biol. Chem.* **2006**, *281*, 578 – 586.
[26] O. Ohlenschläger, T. Seiboth, H. Zengerling, L. Briese, A. Marchanka, R. Ramachandran, M. Baum, M. Korbas, W. Meyer-Klaucke, M. Dürst, M. Görlach, *Oncogene* **2006**, *25*, 5953 – 5959.
[27] L. B. Chemes, I. E. Sanchez, L. G. Alonso, G. de Prat-Gay in *Flexible Viruses: Structural Disorder in Viral Proteins* (Eds.: V. Uversky, S. Longhi), Wiley, Hoboken, **2012**, pp. 313 – 346.
[28] N. Dyson, P. M. Howley, K. Münger, E. Harlow, *Science* **1989**, *243*, 934 – 937.
[29] J. O. Lee, A. A. Russo, N. P. Pavletich, *Nature* **1998**, *391*, 859 – 865.
[30] L. B. Chemes, I. E. Sanchez, C. Smal, G. de Prat-Gay, *FEBS J.* **2010**, *277*, 973 – 988.
[31] K. Huh, X. Zhou, H. Hayakawa, J. Y. Cho, T. A. Libermann, J. Jin, J. W. Harper, K. Münger, *J. Virol.* **2007**, *81*, 9737 – 9747.
[32] E. Berezutskaya, S. Bagchi, *J. Biol. Chem.* **1997**, *272*, 30135 – 30140.
[33] N. Dyson, P. Guida, K. Münger, E. Harlow, *J. Virol.* **1992**, *66*, 6893 – 6902.
[34] R. Davies, R. Hicks, T. Crook, J. Morris, K. Vousden, *J. Virol.* **1993**, *67*, 2521 – 2528.
[35] J. A. Marsh, V. K. Singh, Z. Jia, J. D. Forman-Kay, *Protein Sci.* **2006**, *15*, 2795 – 2804.
[36] J. G. de La Torre, M. L. Huertas, B. Carrasco, *J. Magn. Reson.* **2000**, *147*, 138 – 146.
[37] L. G. Alonso, M. M. Garcia-Alai, A. D. Nadra, A. N. Lapeña, F. L. Almeida, P. Gualfetti, G. de Prat-Gay, *Biochemistry* **2002**, *41*, 10510 – 10518.
[38] L. G. Alonso, M. M. Garcia-Alai, C. Smal, J. M. Centeno, R. Iacono, E. Castaño, P. Gualfetti, G. de Prat-Gay, *Biochemistry* **2004**, *43*, 3310 – 3317.
[39] P. Tompa in *Structure and Function of Intrinsically Disordered Proteins*, CRC Press, Boca Raton, **2009**.
[40] J. M. Firzlaff, B. Lüscher, R. N. Eisenman, *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 5187 – 5191.
[41] B. Todorovic, P. Massimi, K. Hung, G. S. Shaw, L. Banks, J. S. Mymryk, *J. Virol.* **2011**, *85*, 10048 – 10057.
[42] D. R. Patrick, A. Oliff, D. C. Heimbrook, *J. Biol. Chem.* **1994**, *269*, 6842 – 6850.
[43] C. A. Moody, L. A. Laimins, *Nat. Rev. Cancer* **2010**, *10*, 550 – 560.
[44] D. Fera, D. C. Schultz, S. Hodawadekar, M. Reichman, P. S. Donover, J. Melvin, S. Troutman, J. L. Kissil, D. M. Huryn, R. Marmorstein, *Chem. Biol.* **2012**, *19*, 518 – 528.
[45] E. A. Mirecka, R. Rudolph, T. Hey, *Protein Expression Purif.* **2006**, *48*, 281 – 291.
[46] P. Schanda, E. Kupce, B. Brutscher, *J. Biomol. NMR* **2005**, *33*, 199 – 211.
[47] L. E. Kay, P. Keifer, T. Saarinen, *J. Am. Chem. Soc.* **1992**, *114*, 10663 – 10665.
[48] J. Schleucher, M. Schwendinger, M. Sattler, P. Schmidt, O. Schedletzky, S. J. Glaser, O. W. Sørensen, C. Griesinger, *J. Biomol. NMR* **1994**, *4*, 301 – 306.
[49] S. Grzesiek, A. Bax, *J. Magn. Reson.* **1992**, *96*, 432 – 440.
[50] J. Schleucher, M. Sattler, C. Griesinger, *Angew. Chem.* **1993**, *105*, 1518 – 1521; *Angew. Chem. Int. Ed. Engl.* **1993**, *32*, 1489 – 1491.
[51] R. T. Clubb, G. Wagner, *J. Biomol. NMR* **1992**, *2*, 389 – 394.
[52] S. Grzesiek, A. Bax, *J. Am. Chem. Soc.* **1992**, *114*, 6291 – 6293.
[53] D. R. Muhandiram, L. E. Kay, *J. Magn. Reson. Ser. B* **1994**, *103*, 203 – 216.
[54] R. Weisemann, H. Rüterjans, W. Bermel, *J. Biomol. NMR* **1993**, *3*, 113 – 120.
[55] W. Bermel, I. Bertini, L. Duma, L. Emsley, I. C. Felli, R. Pierattelli, P. R. Vasos, *Angew. Chem.* **2005**, *117*, 3149 – 3152; *Angew. Chem. Int. Ed.* **2005**, *44*, 3089 – 3092.
[56] Z. Serber, C. Richter, D. Moskau, J.-M. Boehlen, T. Gerfin, D. Marek, M. Haeberli, L. Baselgia, F. Laukien, A. S. Stern, J. C. Hoch, V. Dötsch, *J. Am. Chem. Soc.* **2000**, *122*, 3554 – 3555.
[57] W. Bermel, I. Bertini, V. Csizmok, I. C. Felli, R. Pierattelli, P. Tompa, *J. Magn. Reson.* **2009**, *198*, 275 – 281.
[58] N. A. Farrow, R. Muhandiram, A. U. Singer, S. M. Pascal, C. M. Kay, G. Gish, S. E. Shoelson, T. Pawson, J. D. Forman-Kay, L. E. Kay, *Biochemistry* **1994**, *33*, 5984 – 6003.
[59] R. Keller, *The Computer Aided Resonance Assignment Tutorial*, Cantina, Goldau, **2004**, pp. 1 – 81.
[60] Sparky 3, T. D. Goddard, D. G. Kneller, University of California, San Francisco, **2000**.

Received: March 22, 2013
Published online on August 12, 2013

CHEMBIOCHEM

Supporting Information

The Heterogeneous Structural Behavior of E7 from HPV16 Revealed by NMR Spectroscopy

Eduardo O. Calçada,[a] Isabella C. Felli,*[a, b] Tomáš Hošek,[a] and Roberta Pierattelli*[a, b]

cbic_201300172_sm_miscellaneous_information.pdf

Table 1. Sequence specific assignment of H^N, N, C´, C^α, C^β in HPV 16 E7 at 298 K. Proton resonances were calibrated with respect to the signal of 2,2-dimethylsilapentane-5-sulfonic acid (DSS). Carbon chemical shifts were referred to external 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS), and to the signal of dioxane at 69.31 ppm as secondary reference. Nitrogen chemical shifts were referenced indirectly to the ^{13}C standard using the conversion factor derived from the ratio of NMR frequencies.[1]

#	Res	HN	N	C´	C^α	C^β
1	MET					
2	HIS					
3	GLY			173.881	45.263	--
4	ASP	8.277	120.44	176.262	54.241	41.051
5	THR	8.159	116.529	172.863	59.819	69.781
6	PRO	--	138.96	177.005	63.336	31.946
7	THR	8.243	115.122	174.411	62.034	69.799
8	LEU	8.206	124.585	176.927	55.419	42.207
9	HIS	8.182	120.201	175.481	56.658	30.77
10	GLU					
11	TYR	8.129	120.234	175.783	58.039	38.417
12	MET	8.128	121.422	175.941	55.522	32.546
13	LEU	7.995	122.589	176.927	55.544	42.362
14	ASP	8.23	120.591	175.831	54.262	41.038
15	LEU	7.972	121.987	177.073	54.837	42.285
16	GLN	8.355	122.313	173.941	53.573	28.73
17	PRO	--	137.091	176.967	63.044	31.954
18	GLU	8.621	120.934	176.871	56.795	30.188
19	THR	8.214	115.068	174.626	61.686	69.784
20	THR	8.159	116.142	174.201	61.977	69.775
21	ASP	8.332	122.949	176.025	54.458	41.046
22	LEU	8.042	122.239	177.147	55.426	42.215
23	TYR	8.128	120.397	175.554	57.969	38.412
24	CYS	7.979	120.913	174.102	58.351	27.603
25	TYR	8.136	123.072	175.791	58.164	38.69
26	GLU	8.269	122.277	175.994	56.601	30.347
27	GLN	8.261	121.354	175.756	55.699	29.312
28	LEU	8.253	123.759	177.055	55.414	42.218
29	ASN	8.41	119.622	174.818	53.065	39.156
30	ASP	8.342	121.476	176.285	54.243	41.328
31	SER	8.3	116.529	174.687	58.347	63.913
32	SER	8.441	118.168	174.622	58.585	63.825
33	GLU	8.449	122.626	176.501	56.582	30.204
34	GLU	8.41	122.095	176.185	56.286	30.459
35	GLU	8.394	121.945	176.489	56.479	30.263
36	ASP	8.355	121.572	176.214	54.245	41.327
37	GLU	8.384	121.632	176.463	56.597	30.201
38	ILE	8.206	122.224	176.022	61.284	38.701
39	ASP	8.428	124.114	176.263	54.245	41.338
40	GLY	8.081	109.191	171.846	44.797	--
41	PRO	--	134.462	177.2	63.053	31.951
42	ALA	8.452	124.265	178.516	52.688	19.049
43	GLY	8.324	107.976	174.064	45.35	--
44	GLN	8.152	119.761	175.552	55.525	29.595
45	ALA	8.355	125.712	177.46	52.355	19.331
46	GLU	8.396	121.677	174.7	56.657	30.199

Table 2. NMR experiments and acquisition parameters used for sequence specific assignment of H^N, N, C', C^α, $C^ß$ resonances and dynamic characterization of HPV16 E7 at 298 K.

The following parameters were used in all experiments. The ^1H carrier was placed at 4.7 ppm, the ^{15}N carrier at 116.9 ppm. Band selective ^{13}C pulses were given at 173 and 39 ppm to excite or invert C' and $C^{\alpha/ß}$ spins respectively. The following band-selective pulses were used: 300µs with Q5[2] and time reversed Q5 shapes for C^α and C^{ali} excitation, 220 µs Q3[2] shape for C'/C^{ali} inversion. The experiments for sequence specific assignment were acquired with a Bruker AVANCE 800 spectrometer equipped with an inverse triple-resonance cryogenically cooled probehead (^1H-^{15}N SOFAST-HMQC[3], HNCO,[4;5] HNCACO,[6] HNCACB,[7] CBCACONH,[8] HN(CA)NNH[9]) and with a Bruker AVANCE 700 spectrometer equipped with a triple-resonance cryogenically cooled probehead optimized for direct ^{13}C-detection (CON-IPAP,[10] HCACO-IPAP,[11;12] HCBCACON-IPAP[12]). The other experiments (^{15}N R_1, ^{15}N R_2 and ^1H-^{15}N NOEs) were acquired with a Bruker AVANCE 500 spectrometer equipped with an inverse triple-resonance cryogenically cooled probehead.

Experiments	Dimension of acquired data			Spectral width (ppm)			n[a]	d[b]
	t_1	t_2	t_3	F_1	F_2	F_3		
^{13}C detected								
CON-IPAP[c]	900* (^{15}N)	1024 (^{13}C)		35	30		32	2.0
HCACO-IPAP[c]	660* (^{13}C)	1024 (^{13}C)		50	32		32	1.4
HCBCACON-IPAP[c]	220* (^{13}C)	64 (^{15}N)	1024 (^{13}C)	60.4	36	50	16	1.2
^1H detected								
^1H-^{15}N SOFAST-HMQC	512 (^{15}N)	1024 (^1H)		22	16		2	0.5
HNCO	128 (^{13}C)	60 (^{15}N)	2048 (^1H)	10	22	13.9	4	1
HNCACO	128 (^{13}C)	60 (^{15}N)	2048 (^1H)	10	22	13.9	16	1
CBCACONH	128 (^{13}C)	60 (^{15}N)	2048 (^1H)	75	22	13.9	8	1
HNCACB	128 (^{13}C)	60 (^{15}N)	2048 (^1H)	75	22	13.9	16	1
HN(CA)NNH	70 (^{15}N)	70 (^{15}N)	2048 (^1H)	25	25	13.9	32	1
Dynamics								
^{15}N T_1[d]	128 (^{15}N)	1024 (^1H)		25	14		32	3.0
^{15}N T_2[d]	128 (^{15}N)	1024 (^1H)		25	14		32	3.0
steady-state heteronuclear ^1H-^{15}N-NOEs[d]	256 (^{15}N)	1024 (^1H)		25	14		168	5.0

[a] number of acquired scans.

[b] Relaxation delay in seconds.

[c] For experiments acquired in the IPAP mode, the dimension in which the two experiments are stored is indicated with an asterisk.

[d] In ^{15}N R_1, ^{15}N R_2 and heteronuclear ^1H-^{15}N-NOEs experiments the water signal was suppressed with 'water flip-back' scheme.[13] For the determination of R_1, 10 experiments were acquired changing the variable delay from 2.5 ms to 1.2 s. For the determination of R_2, 9 experiments were acquired changing the variable delay from 17 to 250 ms.

All 3D and 2D spectra were processed using TopSpin 2.0 (Bruker) and analyzed through the CARA (Computer Aided Resonance Assignment) software developed at ETH-Zürich.

Figure S1. The 2D ^{1}H-^{15}N correlation spectrum of the short construct comprising residues 45-98 (E7C) is shown. The spectrum was recorded on a Bruker AVANCE NMR spectrometer operating at 900 MHz and 298 K on a 0.04 mM sample in 10 mM HEPES buffer at pH 7.5, 150 mM KCl, 10 μM $ZnCl_2$ and 10 mM DTT.

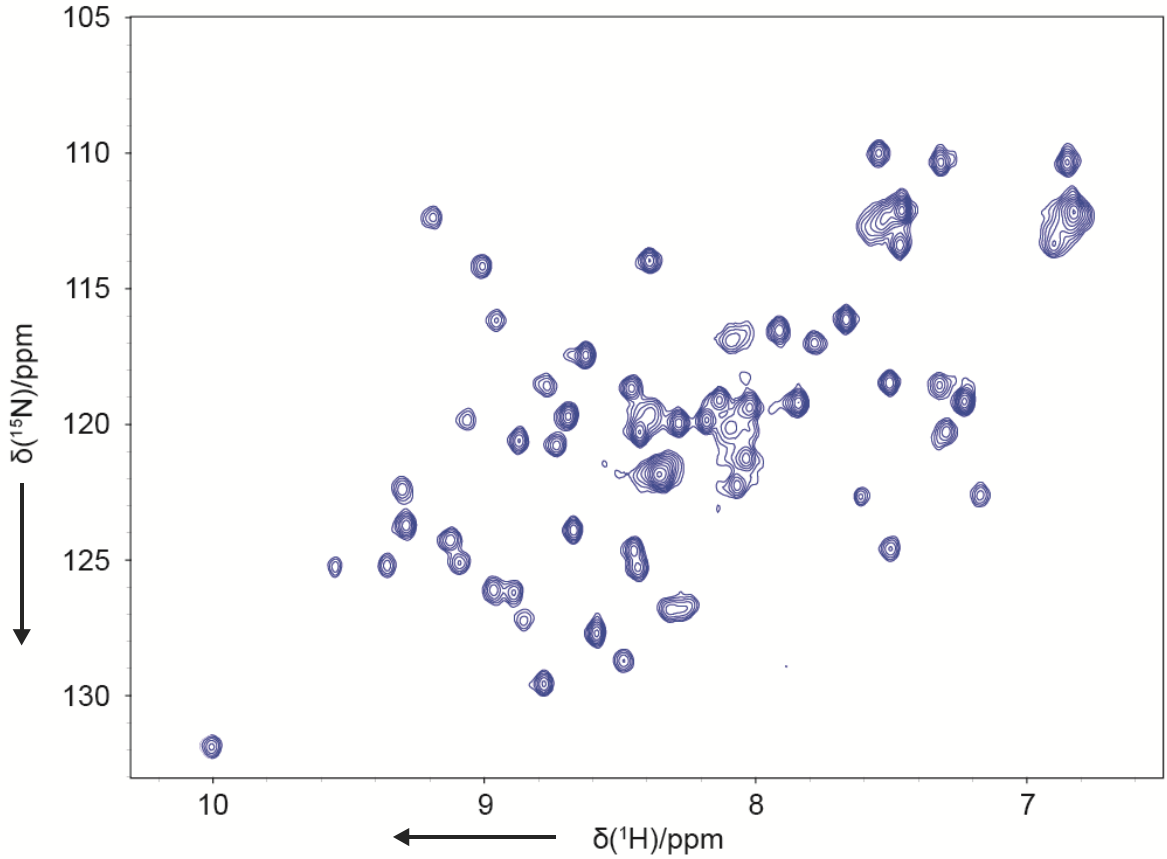
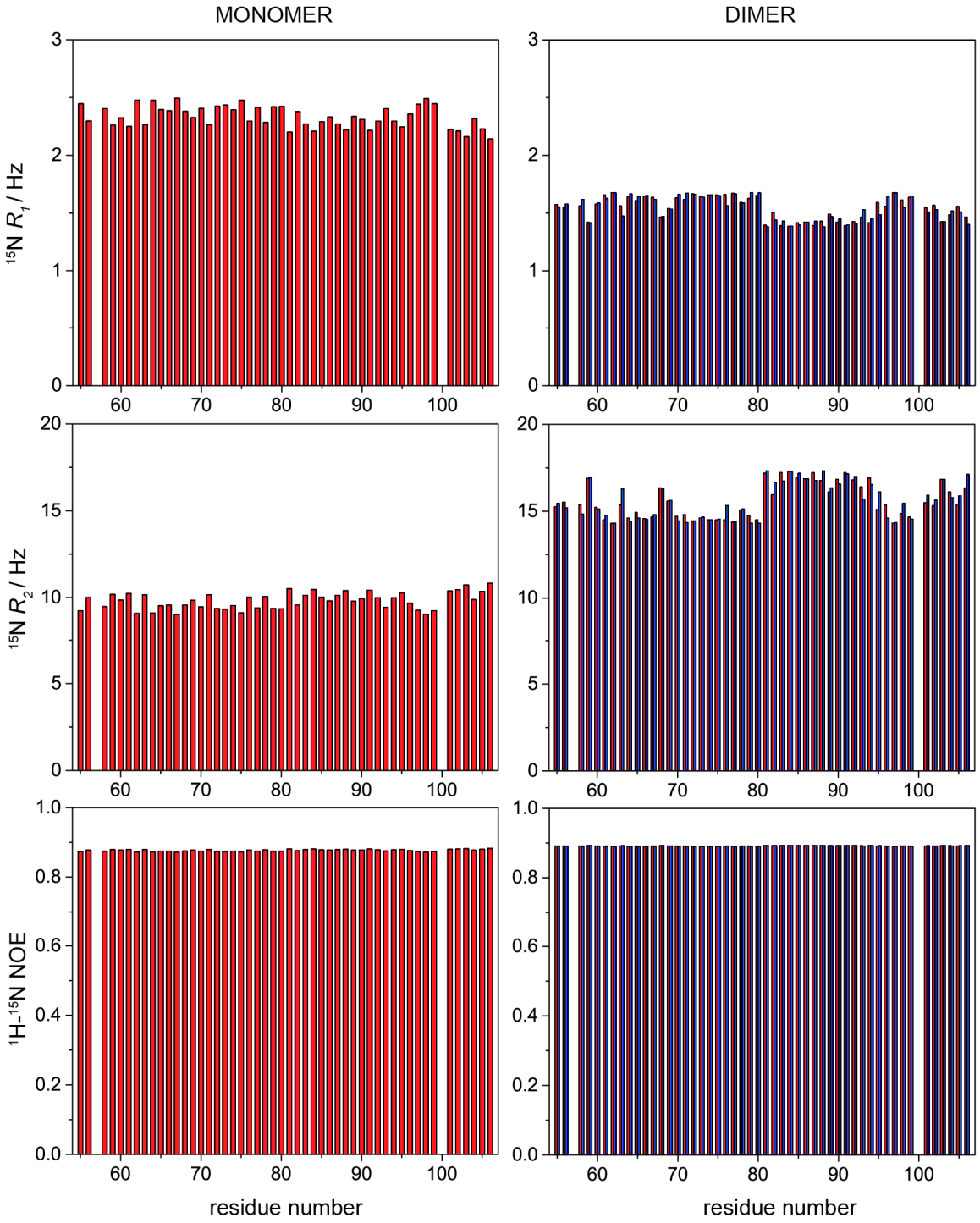


Figure S2. The ^{15}N relaxation data (^{15}N R_1, ^{15}N R_2 and ^{1}H-^{15}N NOE) calculated with HYDRONMR[14] using as input models the monomeric and dimeric NMR structures of the HPV 45 E7 CR3 (PDB codes 2EWL and 2F8B)[15] are reported below. The data calculated for the monomer are reported in the left panels and those for the dimer in the right ones.



Reference List

[1.] J. L. Markley, A. Bax, Y. Arata, C. W. Hilbers, R. Kaptein, B. D. Sykes, P. E. Wright, K. Wüthrich, *J.Biomol.NMR* **1998,** *12* 1-23.

[2.] L. Emsley, G. Bodenhausen, *Chem.Phys.Lett.* **1990,** *165* 469-476.

[3.] P. Schanda, B. Brutscher, *J.Am.Chem.Soc.* **2005,** *127* 8014-8015.

[4.] S. Grzesiek, A. Bax, *J.Magn.Reson.* **1992,** *96* 432-440.

[5.] J. Schleucher, M. Sattler, C. Griesinger, *Angew.Chem.Int.Ed Engl.* **1993,** *32* 1489-1491.

[6.] R. T. Clubb, G. Wagner, *J.Biomol.NMR* **1992,** *2* 389-394.

[7.] D. R. Muhandiram, L. E. Kay, *J.Magn.Reson.Ser.B* **1994,** *103* 203-216.

[8.] S. Grzesiek, A. Bax, *J.Am.Chem.Soc.* **1992,** *114* 6291-6293.

[9.] R. Weisemann, H. Rüterjans, W. Bermel, *J.Biomol.NMR* **1993,** *3* 113-120.

[10.] W. Bermel, I. Bertini, L. Duma, L. Emsley, I. C. Felli, R. Pierattelli, P. R. Vasos, *Angew.Chem.Int.Ed.* **2005,** *44* 3089-3092.

[11.] Z. Serber, C. Richter, D. Moskau, J.-M. Boehlen, T. Gerfin, D. Marek, M. Haeberli, L. Baselgia, F. Laukien, A. S. Stern, J. C. Hoch, V. Dötsch, *J.Am.Chem.Soc.* **2000,** *122* 3554-3555.

[12.] W. Bermel, I. Bertini, V. Csizmok, I. C. Felli, R. Pierattelli, P. Tompa, *J.Magn.Reson.* **2009,** *198* 275-281.

[13.] S. Grzesiek, A. Bax, *J.Am.Chem.Soc.* **1993,** *115* 12593-12594.

[14.] J. G. de la Torre, M. L. Huertas, B. Carrasco, *J.Magn.Reson.* **2000,** *147* 138-146.

[15.] O. Ohlenschlager, T. Seiboth, H. Zengerling, L. Briese, A. Marchanka, R. Ramachandran, M. Baum, M. Korbas, W. Meyer-Klaucke, M. Dürst, M. Görlach, *Oncogene* **2006,** *25* 5953-5959.

5.2.2

Biophysical characterization of the E7 conserved region 3 from HPV 16

Tomáš Hošek[a], Eduardo O. Calçada[a], Marcela O. Nogueira[a], Isabella C. Felli[a], Roberta Pierattelli[a]

[a] *CERM and Department of Chemistry Ugo Schiff, University of Florence, via Luigi Sacconi 6, Sesto Fiorentino, 50019, Italy*

In preparation

Introduction

Many intrinsically disordered proteins (IDPs) consist of combinations of globular domains, linear motifs and disordered regions. These IDPs are characterized by a dynamic ensemble of conformations that cannot be easily described by traditional structural biology and have often a heterogeneous structural behavior that render their investigation troublesome.

This is the case of the protein E7 from Human Papilloma Virus (HPV). This protein has been investigated since long time as it is the main transforming product from HPV [1] and has been reported to interact with a large number of cellular targets. This feature enables E7 to interfere with multiple cell regulatory pathways, eventually leading to cell transformation and cancer in the case of the most aggressive HPV phenotypes. [2]

E7 is constituted by about one hundred amino acids that can be grouped in three conserved regions (CR), namely CR1 and CR2 in the N-terminal half (E7N) and CR3 in the C-terminal half (E7C). [3] The first two conserved regions (CR1 and CR2) show a high propensity to be disordered. The CR3 region, which contains two CXXC motifs separated by 29 or 30 amino acids, has been shown to bind zinc and is more structured. [4] The structural characterization of the C-terminal part of the protein, which contains the CR3 and the zinc binding motifs, has been achieved for short constructs of the HPV 1 variant through X-ray crystallography [5] and of the HPV 45 variant through nuclear magnetic resonance (NMR) spectroscopy [6]. The E7 CR3 of these variants assembles as a dimer. Each subunit of the dimer contains a two-stranded antiparallel sheet formed by residues 4452 and 5865 followed by a sharp U-turn leading to 12 amino acids long helix, a bend and an extended strand, leading to a final short helix. [5] If the CR3 region is structurally well characterized, the whole protein failed to give crystals and was not characterized through NMR until very recently. [7]

In the previous work [7] we have studied with high-resolution techniques for the first time the full construct of E7 from HPV 16, one of the most relevant variant of the virus, which proved to be responsible for the occurrence of cervix cancer. [8] This protein shares most of the characteristics of E7 isolated from other variants, but resulted to be even more difficult to be investigated due to the high tendency to form macromolecular assemblies. [9] We have shown that the combined use of ^{1}H and ^{13}C direct detected NMR experiments on the entire construct of HPV 16 E7 can provide the complete assignment of the first half of the protein, which showed the high mobility and the favorable relaxation properties of an IDP [7] while the second part failed to give useful triple resonance spectra due to large signal broadening and poor sensitivity. The latter effect is likely the result of

both the tendency of this protein to form large molecular assemblies and the chemical exchange between monomeric/dimeric/tetrameric forms of the protein (and higher order multimers). [10, 11] The formation of these multimers is mediated by the CR3 region of HPV E7 but mutations that prevent E7 oligomerization lead to protein instability and disrupt the cell transforming properties of E7 [12] render the investigation of these mutated forms of the protein far less significant than studies of the wild-type protein. However, the characterization of the CR3 is crucial to understand the factor modulating the interaction of HPV 16 E7 with the retinoblastoma binding protein, which is known to be degradeted by the concerted action of HPV 16 E7 CR2 and CR3 regions. [13]

We thus decide to undertake the characterization of the isolated CR3 domain of the protein and be able to establish the existence of interactions between the IDP part of the protein and the folded one, eventually responsible for the peculiar dynamic properties of the HPV 16 variant.

Materials and methods

Protein preparation and purification

The recombinant vector pET21a-E7-CR3 was transformed into *E. coli* BL21(DE3)pLysS. After expression tests, one colony was selected to grow in LB medium with ampicillin and chloramphenicol. Finally, the cells were transferred to 1 l of ^{15}N or ^{13}C, ^{15}N minimal medium with ampicillin, chloramphenicol and 100 μM of $ZnCl_2$ and allowed to grow at 37 °C. When the optical density at 600 nm reached 0.6, the cells were induced with 1 mM of IPTG and were allowed to grow for an additional 4 hours at 37 °C. The E7-CR3 domain, expressed in the inclusion bodies, was re-suspended in 1 M and 2 M urea. The solutions containing the protein were mixed, concentrated and injected in a size exclusion column (Superdex 75 16/100) pre-equilibrated with 50 mM Tris pH 8.5, 50 mM KCl, 10 mM DTT and 10 μM $ZnCl_2$. The fractions containing the enriched protein (either ^{15}N or ^{13}C, ^{15}N) were mixed and concentrated to 0.1 mM for the NMR experiments.

NMR Experiments

The ^{1}H detected NMR experiments for sequence specific assignment were acquired at 298 K on the 22.3 T Bruker AVANCE 950 spectrometer, operating at 950.2 MHz for ^{1}H, equipped with cryogenically cooled TCI probe. A datasets of ^{1}H detected NMR experiments (^{1}H-^{15}N BEST-TROSY [14, 15], ^{1}H-^{13}C HSQC [16–18], BEST-TROSY HNCO

[19–21], BEST-TROSY HNCA [19–21], CBCACONH [22], HCCH-TOCSY [23]) were acquired for sequence specific assignment of H^N, N, C', C^α, C^β resonances. To characterize the structural and dynamic features of HPV 16 E7, ^{15}N relaxation experiments (R_1, R_2 and ^{1}H-^{15}N NOEs) on a 11.8 T Bruker AVANCE 500 spectrometer equipped with a cryogenically cooled probe. All experimental parameters used to acquire NMR experiments described above are reported in more detail in Table 2. Data were processed with TopSpin 2.0 and were analyzed with program ccpNMR [24]. The secondary structure propensity from chemical shifts was determined using the ncSPC web program [www. http://nmr.chem.rug.nl/ncSPC/]. [25] The ^{15}N relaxation rates (R_1 and R_2) were determined by fitting the cross-peak intensity measured as a function of variable delay, to single-exponential decay. ^{1}H-^{15}N NOE values were obtained as a ratio between peak intensity in spectra recorded with and without ^{1}H saturation.

Results and Discussions

The tendency of the CR3 region to form larger molecular assemblies and the exchange between several multimeric forms has a negative effect on the NMR relaxation properties, which limits the choice of the experiments for the sequence specific assignment. Hence, the set of the most sensitive 3D experiments were selected for our investigation. The analysis of the acquired spectra allowed us to assign 68% of the polypeptide chain. The sequence specific assignment is indicated in the 2D ^{1}H-^{15}N correlated spectrum acquired on CR3 construct of HPV 16 E7 protein (Figure 1). Chemical shifts of $^1H^N$, ^{15}N, ^{13}C', $^{13}C^\alpha$ and $^{13}C^\beta$ of the assigned residues are reported in Table 1.

The obtained data from the sequence specific assignment were used for the SSP analysis. In Figure 2 the SSP score for the initial disordered N-terminal half of E7 is also shown. [7] The results are in agreement with the structural investigations on the other HPV serotypes [5, 6], and confirm the high propensity of the initial N terminal part of CR3 (residue 45-58) to adopt extended conformation and the presence of two α-helixes occurring in regions from residue 70 to 80 and 91 to 95. The part of the CR3 that is reported to be on the interface of the dimeric form, were not sequentially assigned due to unambiguous and missing data. [26]

Nevertheless, the partial sequence specific assignment of the isolated CR3 region allowed us to transfer assignment on the 2D ^{1}H-^{15}N correlated spectrum of the full-length E7, which enable us to analyze the ^{15}N relaxation data obtained on the full-leght protein with the corresponding residues of the CR3.

The ^{15}N R_1, ^{15}N R_2 relaxation rates and ^{1}H-^{15}N heteronuclear NOEs as the function of the residue number of the whole E7 protein are reported in Figure 3. The data obtained for the CR3 signals are affected by large errors, but clearly pinpoint the presence of a heterogeneous behavior of these nuclei, suggesting that structure of the CR3 of E7 from the HPV 16 is not as well-defined as it is for other serotypes studied by high resolution techniques.

Further work is in progress in order to complete characterization of structural and dynamical features of this domain.

References

[1] Phelps WC, Yee CL, Munger K, Howley PM. The human papillomavirus type 16 E7 gene encodes transactivation and transformation functions similar to those of adenovirus E1A. *Cell*. 1988;53(4):539–547.

[2] Moody CA, Laimins LA. Human papillomavirus oncoproteins: pathways to transformation. *Nat Rev Cancer*. 2010;10(8):550–560.

[3] Phelps WC, Munger K, Yee CL, Barnes JA, Howley PM. Structure-function analysis of the human papillomavirus type 16 E7 oncoprotein. *Journal of Virology*. 1992; 66(4):2418–2427.

[4] Uversky VN, Roman A, Oldfield CJ, Dunker AK. Protein Intrinsic Disorder and Human Papillomaviruses: Increased Amount of Disorder in E6 and E7 Oncoproteins from High Risk HPVs. *J Proteome Res*. 2006;5(8):1829–1842.

[5] Liu X, Clements A, Zhao K, Marmorstein R. Structure of the human Papillomavirus E7 oncoprotein and its mechanism for inactivation of the retinoblastoma tumor suppressor. *J Biol Chem*. 2006;281(1):578–586.

[6] Ohlenschläger O, Seiboth T, Zengerling H, Briese L, Marchanka A, Ramachandran R, Baum M, Korbas M, Meyer-Klaucke W, Dürst M, Görlach M. Solution structure of the partially folded high-risk human papilloma virus 45 oncoprotein E7. *Oncogene*. 2006;25(44):5953–5959.

[7] Calçada EO, Felli IC, Hošek T, Pierattelli R. The Heterogeneous Structural Behavior of E7 from HPV16 Revealed by NMR Spectroscopy. *ChemBioChem*. 2013; 14(14):1876–1882.

[8] zur Hausen H. Papillomaviruses and cancer: from basic studies to clinical application. *Nat Rev Cancer*. 2002;2(5):342–350.

[9] Clements A, Johnston K, Mazzarelli JM, Ricciardi RP, Marmorstein R. Oligomerization Properties of the Viral Oncoproteins Adenovirus E1A and Human Papillomavirus E7 and Their Complexes with the Retinoblastoma Protein. *Biochemistry*. 2000;39(51):16033–16045.

[10] Chinami M, Sasaki S, Hachiya N, Yuge K, Ohsugi T, Maeda H, Shingu M. Functional oligomerization of purified human papillomavirus types 16 and 6b E7 proteins expressed in Escherichia coli. *J Gen Virol*. 1994;75 (Pt 2)(2):277–281.

[11] Munger K, Howley P, DiMaio D. Human Papillomavirus E6 and E7 Oncogenes. In: *The Papillomaviruses*, pp. 197–252. Boston, MA: Springer US. 2007;.

[12] Fera D, Marmorstein R. Different Regions of the HPV-E7 and Ad-E1A Viral Oncoproteins Bind Competitively but through Distinct Mechanisms to the CH1 Transactivation Domain of p300. *Biochemistry*. 2012;51(47):9524–9534.

[13] Todorovic B, Hung K, Massimi P, Avvakumov N, Dick FA, Shaw GS, Banks L, Mymryk JS. Conserved region 3 of human papillomavirus 16 E7 contributes to deregulation of the retinoblastoma tumor suppressor. *Journal of Virology*. 2012;86(24):13313–13323.

[14] Favier A, Brutscher B. Recovering lost magnetization: polarization enhancement in biomolecular NMR. *J Biomol NMR*. 2011;49(1):9–15.

[15] Lescop E, Kern T, Brutscher B. Guidelines for the use of band-selective radiofrequency pulses in hetero-nuclear NMR: example of longitudinal-relaxation-enhanced BEST-type 1H-15N correlation experiments. *J Magn Reson*. 2010;203(1):190–198.

[16] Palmer AG III, Cavanagh J, Wright PE, Rance M. Sensitivity improvement in proton-detected two-dimensional heteronuclear correlation NMR spectroscopy. *J Magn Reson (1969-1992)*. 1991;93(1):151–170.

[17] Kay LE, Keifer P, Saarinen T. Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J Am Chem Soc*. 1992;114(26):10663–10665.

[18] Schleucher J, Schwendinger M, Sattler M, Schmidt P, Schedletzky O, Glaser SJ, Sørensen OW, Griesinger C. A general enhancement scheme in heteronuclear multidimensional NMR employing pulsed field gradients. *J Biomol NMR*. 1994;4(2):301–306.

[19] Lescop E, Schanda P, Brutscher B. A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *Journal of Magnetic Resonance*. 2007; 187(1):163–169.

[20] Salzmann MM, Pervushin K, Wider GG, Senn HH, Wüthrich K. TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. *Proc Natl Acad Sci USA*. 1998;95(23):13585–13590.

[21] Solyom Z, Schwarten M, Geist L, Konrat R, Willbold D, Brutscher B. BEST-TROSY experiments for time-efficient sequential resonance assignment of large disordered proteins. *J Biomol NMR*. 2013;55(4):311–321.

[22] Grzesiek S, Bax A. Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J Am Chem Soc*. 1992; 114(16):6291–6293.

[23] Kay LE, Xu GY, Singer AU, Muhandiram DR, Forman-Kay JD. A Gradient-Enhanced HCCH-TOCSY Experiment for Recording Side-Chain1H and13C Correlations in H2O Samples of Proteins. *J Magn Reson B*. 1993;101(3):333–337.

[24] Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J, Laue ED. The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins*. 2005;59(4):687–696.

[25] Tamiola K, Mulder FAA. Using NMR chemical shifts to calculate the propensity for structural order and disorder in proteins. *Biochemical Society transactions*. 2012; 40(5):1014–1020.

[26] Todorovic B, Massimi P, Hung K, Shaw GS, Banks L, Mymryk JS. Systematic Analysis of the Amino Acid Residues of Human Papillomavirus Type 16 E7 Conserved Region 3 Involved in Dimerization and Transformation. *Journal of Virology*. 2011; 85(19):10048–10057.

[27] Markley JL, Bax A, Arata Y, Hilbers CW, Kaptein R, Sykes BD, Wright PE, Wüthrich K. Recommendations for the presentation of NMR structures of proteins

and nucleic acids–IUPAC-IUBMB-IUPAB Inter-Union Task Group on the standardization of data bases of protein and nucleic acid structures determined by NMR spectroscopy. *Eur J Biochem.* 1998;256(1):1–15.

[28] Grzesiek S, Bax A. The importance of not saturating water in protein NMR. Application to sensitivity enhancement and NOE measurements. *J Am Chem Soc.* 1993; 115(26):12593–12594.

Figures

Figure 1. ^{1}H-^{15}N BEST-TROSY spectra recorded for CR3 construct of the E7 protein of HPV 16. The resulting assignment is indicated in the figure. The experiment was recorded at 22.3 T spectrometer at 298 K for a 0.1 mM sample in 50 mM Tris, pH 8.5, 50 mM KCl, 10 μM $ZnCl_2$ and 10 mM DTT.

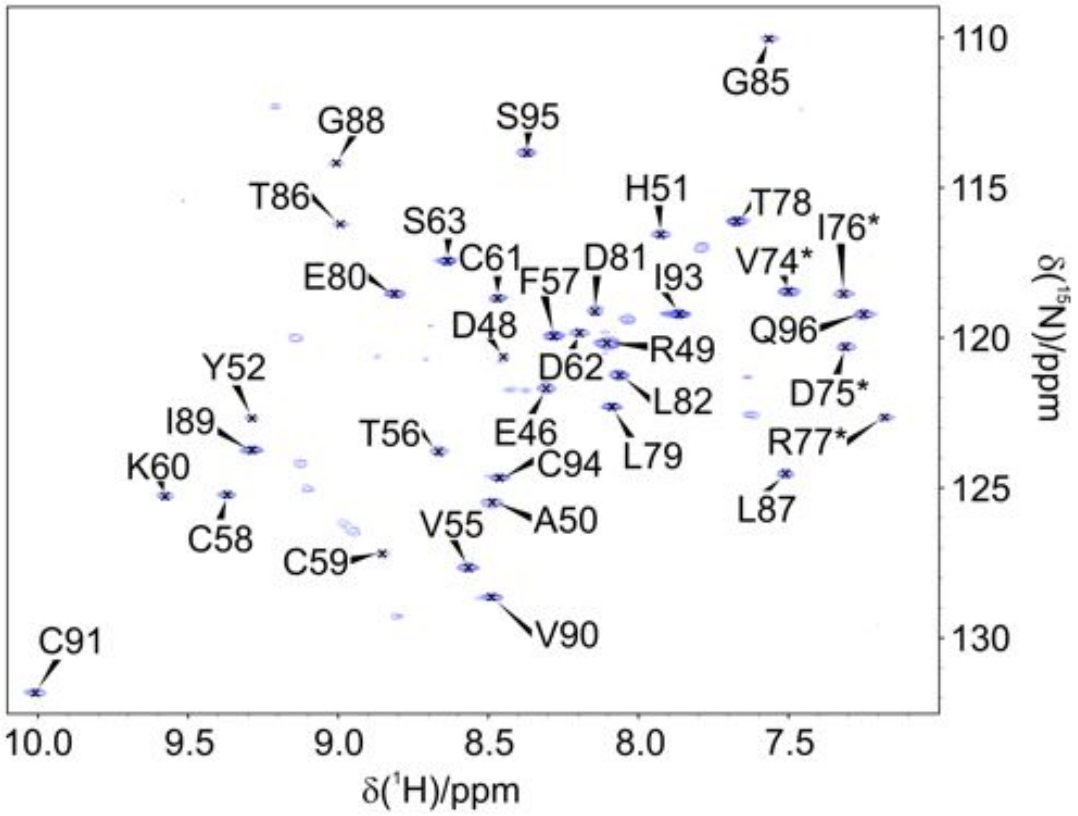


Figure 2. Secondary structure propensity (SSP) [25] score obtained by combining N, C', C^{α} and C^{β} chemical shift for assigned resonances of whole HPV 16 E7 protein (data for N-terminal half were taken from [7]).

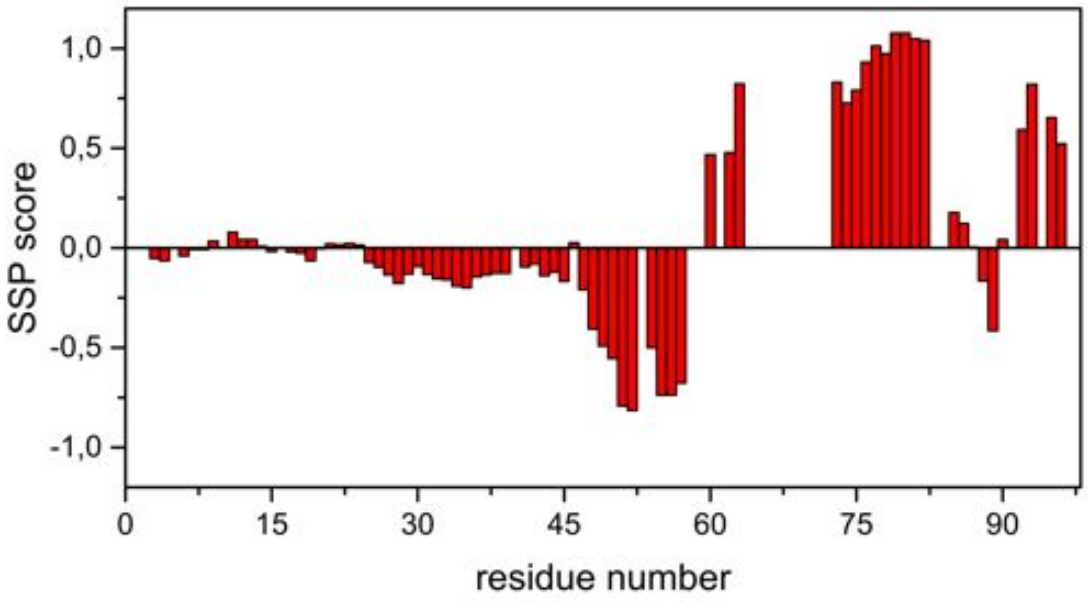


Figure 3. The relaxation data for backbone ^{15}N nuclei as a function of the residue number. From top to bottom: ^{15}N R_1 rate, ^{15}N R_2 rate and ^{1}H-^{15}N NOE values. Relaxation data were taken from [7]. Experiments were recorded at 11.8 T spectrometer for a 0.25 mM sample of full-length HPV 16 E7 protein in 20 mM HEPES buffer at pH 7.5, 150 mM KCl, 10 μM $ZnCl_2$ and 10 mM DTT.

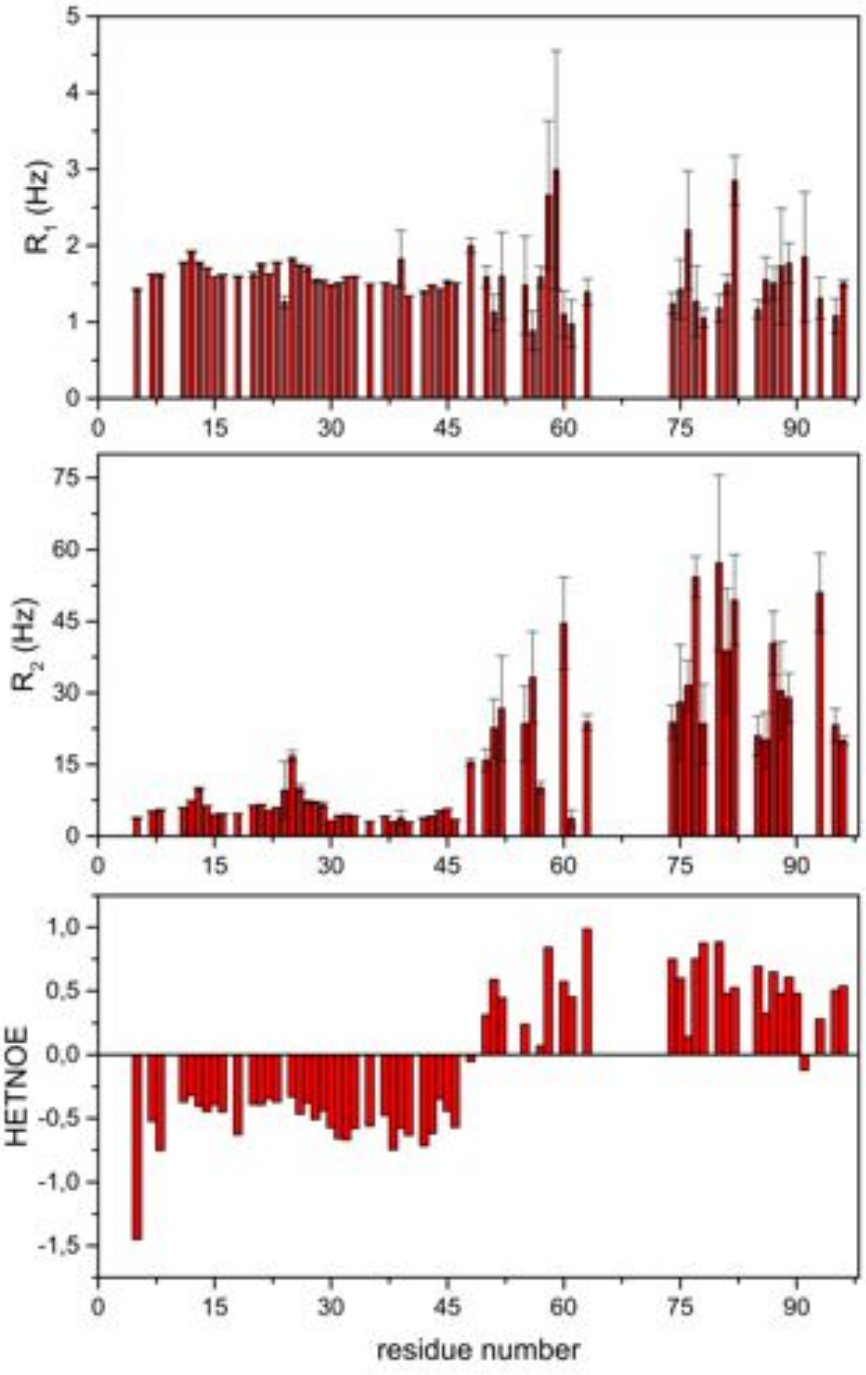


Table 1: Sequence specific assignment of H^N, N, C', C^α and C^β in CR3 construct from HPV 16 E7 protein. Proton resonances were calibrated with respect to the signal of 2,2-dimethylsilapentane-5-sulfonic acid (DSS). Carbon and nitrogen resonances were referenced indirectly to the ^{1}H standard using the conversion factor derived from the ratio of NMR frequencies. [27] The residue numbers of the full-length E7 were used, ambiguously assigned residues are marked with asterisk.

residue n°	aa type	H^N	N	C'	C^α	C^β
45	A			173.93	52.30	19.27
46	E	8.33	121.70		54.33	
47	P			176.42	63.10	31.96
48	D	8.46	120.69	175.90	54.57	41.04
49	R	8.11	120.22	174.79	55.21	32.63
50	A	8.50	125.49	174.73	50.82	21.22
51	H	7.93	116.60	175.29	55.88	32.51
52	Y	9.28	122.75		57.52	40.04
54	I			174.77	59.94	
55	V	8.57	127.64	175.27	61.84	
56	T	8.67	123.83	170.94	58.43	69.48
57	F	8.29	119.96	175.15	55.37	41.77
58	C	9.37	125.21	178.16	59.13	
59	C	8.85	127.14	174.05	59.73	27.54
60	K	9.58	125.27	177.30	57.61	33.81
61	C	8.47	118.71	176.08	59.16	
62	D	8.20	119.85	175.26	56.93	40.71
63	S	8.64	117.41		60.24	
73	H*			177.15	59.41	30.22
74	V*	7.50	118.47	178.25	65.49	
75	D*	7.31	120.29	177.98	57.25	
76	I*	7.31	118.54	176.71	60.82	
77	R*	7.17	122.70	178.55	58.27	29.21
78	T	7.68	116.15	176.38	66.40	67.98
79	L	8.09	122.28	177.85	58.33	41.62
80	E	8.81	118.52	179.32	60.67	29.65
81	D	8.15	119.13	180.13	57.63	39.99

82	L	8.07	121.24		58.10	
84	M		176.45			
85	G	7.57	110.06	174.72	45.33	
86	T	8.99	116.23	173.88	61.91	68.96
87	L	7.51	124.51	174.52	55.90	
88	G	9.01	114.18	171.82	44.08	
89	I	9.29	123.76	174.21	59.98	
90	V	8.49	128.63	175.72	60.88	
91	C	10.01	131.87		58.09	
92	P			178.98	65.47	32.13
93	I	7.87	119.22	179.32	63.98	37.30
94	C	8.47	124.62	178.05	64.19	29.36
95	S	8.38	113.84	174.49	60.39	63.40
96	Q	7.25	119.20		55.80	

Table 2: The NMR experiments and acquisition parameters used for sequence specific assignment of H^N, N, C', C^α and C^β resonances of CR3 from HPV 16 E7 at 298 K.
The ^1H detected experiments for sequence specific assignment were acquired with a Bruker AVANCE 950 spectrometer equipped with an inverse triple-resonance cryogenically cooled probehead (^1H-^{15}N BEST-TROSY [14, 15], ^1H-^{13}C HSQC [16–18], BEST-TROSY HNCO [19–21], BEST-TROSY HNCA [19–21], CBCACONH [22], HCCH-TOCSY [23]). The ^1H carrier was placed at 4.7 ppm, the ^{15}N carrier at 121.3 ppm and the ^{13}C carrier at 173 ppm for C' and at 39 ppm for C_{ali}. Band selective ^1H pulses were given at 8.9 ppm to excite or invert H^N spins. The following band-selective pulses were used: 1800 μs with Pc9 shape for H^N excitation, 1270 μs EBURP and time reversed EBURP shapes for H^N excitation, 1250 μs REBURP shape for H^N inversion. Band selective ^{13}C pulses were given at 173 and 39 ppm to excite or invert C' and C_{ali} spins respectively. The following band-selective pulses were used: 274 μs with G4 and time reversed G4 shapes for C' and C_{ali} excitation, 190 μs Q3 shape for C'/C_{ali} inversion. The other experiments (^{15}N R_1, ^{15}N R_2 and ^1H-^{15}N NOEs) were acquired with a Bruker AVANCE 500 spectrometer equipped with an inverse triple-resonance cryogenically cooled probehead.

Experiments	**Dimension of acquired data**			**Spectral width (ppm)**			n[a]	d[b]
	t_1	t_2	t_3	F_1	F_2	F_3		
^{1}H detected								
^{1}H-^{15}N BEST-TROSY	256(^{15}N)	1024(^{1}H)		25	12		16	0.2
^{1}H-^{13}C HSQC	640(^{13}C)	1600(^{1}H)		75	12		16	1
BEST-TROSY HNCO	72(^{13}C)	32(^{15}N)	1024(^{1}H)	12	25	12	128	0.25
BEST-TROSY HNCA	68(^{13}C)	32(^{15}N)	1024(^{1}H)	28	25	12	176	0.25
CBCACONH	96(^{13}C)	32(^{15}N)	2048(^{1}H)	70	25	16	108	1
HCCH-TOCSY	128(^{13}C)	128(^{13}C)	2048(^{1}H)	75	27	14	16	1
Structure and dynamics								
^{15}N R_1[c]	128(^{15}N)	1024(^{1}H)		25	14		32	3.0
^{15}N R_2[c]	128(^{15}N)	1024(^{1}H)		25	14		32	3.0
Steady-state heteronuclear ^{1}H-^{15}N NOEs[c]	256(^{15}N)	2048(^{1}H)		25	14		168	5.0

a Number of acquired scans.

b Relaxation delay in seconds.

c In ^{15}N R_1, ^{15}N R_2 and heteronuclear ^{1}H-^{15}N-NOEs experiments the water signal was suppressed with "water flip-back" scheme. [28] For the determination of R_1, 10 experiments were acquired changing the variable delay from 2.5 ms to 1.2 ms. For the determination of R_2, 9 experiments were acquired changing the variable delay from 17 ms to 250 ms. All 3D and 2D spectra were processed using the standard Bruker software (TopSpin 2.0) and analyzed through the CARA (Computer Aided Resonance Assignment) developed at ETH-Zürich.

5.2.3

The highly flexible and heterogeneous nature of HAdV E1A characterized at atomic resolution through NMR

Tomáš Hošek[a], Eduardo O. Calçada[a], Michele Salvi[a], Talita Pagani[a], Isabella C. Felli[a], Roberta Pierattelli[a]

[a] *CERM and Department of Chemistry Ugo Schiff, University of Florence, via Luigi Sacconi 6, Sesto Fiorentino, 50019, Italy*

In preparation

Abstract

The small-DNA Adenovirus encode one of the most versatile molecular hub, the early region E1 (E1A) protein. This protein is essential for productive viral infection in human cells and a vast amount of biologically relevant data are available on its interactions with host proteins. Up to now, however, no high-resolution information on any E1A protein is available, despite its important biological role.

Among the different spliced variants of E1A, two are expressed at high level in the early stage of infection. These are the 243 residues long 12S isoform (E1A-12S) and the 289 residue long 13S isoform (E1A-13S). Here we present their NMR characterization, showing that they are both highly disordered, but also show a certain heterogeneous behavior in terms of structural and dynamic properties. This study opens the way to characterize the many interactions in which this protein is involved.

Running title

NMR studies on HAdV E1A

Keywords

NMR; IDP; viral protein; adenovirus; SLiMs

Introduction

It is well known that the structural and dynamic properties of proteins are an essential aspect determining their function. [1–3] What is less known is all the modes through which highly disordered and flexible proteins or protein modules have an impact on protein function as, until recently, it was believed that only proteins characterized by well folded structures could be functional. [4, 5] However, recent evidence accumulating at a very fast speed shows that also in the absence of a stable 3D structure proteins are functional and that actually a high extent of disorder and flexibility can provide functional advantages in a variety of different situations. [6]

The possibility to adopt different conformations may enable a protein to interact with many different partners. [7–9] This contributes to expanding our view on how proteins interact with each other based on many examples well documented in the PDB, such as with the presence of large complementary surfaces mediating protein-protein interactions. It has been suggested that short linear motifs (SLiMs), in other words short sequences of amino acids with specific patterns, may be sufficient to mediate protein-protein interactions particularly in intrinsically disordered proteins (IDPs). [10] Understanding at the molecular level the mode of interaction driven by SLiMs in IDPs is a theme under the spotlight in modern structural biology, with numerous examples appearing in the literature. [11, 12]

Viruses, with their small genomes, do need economic ways to interfere with host proteins. Therefore exploiting SLiMs in IDPs is an efficient strategy towards the goal of implementing many functions into just a single protein. Moreover, as IDPs do not poses any structural constrains derived from the well-defined 3D structure, alternative splicing can be freely exploited to alter the set of protein functions coded into a single gene and thus maximize the usage of a viral genome. [13]

Human Adenovirous (HAdV) is a well characterized virus, routinely used in molecular biology and a lot of information is available on it. [14–17] Among the few proteins encoded by HAdV we are here focusing on the two major forms of E1A proteins expressed at high levels in the initial stage of a viral infection. [18–20] Indeed, despite all the information available on this virus and all the documented interactions of these proteins with key players in host cells, there is no high-resolution structural and dynamic information available on the whole protein [21], as the 3D structure of the protein could not be resolved by X-ray crystallography. Both proteins arise from alternative splicing of the same transcript of the E1A gene and differ only by the presence of the internal region of 46 residues in the

larger protein. The proteins are referred as 13S (E1A-13S) and 12S (E1A-12S) reflecting the sedimentation coefficient of their respective mRNA. [22]

Therefore we decided to undertake the NMR investigation to the two major forms of E1A in an attempt of understanding the molecular basis of how such a hub protein function.

Materials and methods

Cloning, expression and purification of 12S isoform of E1A protein from HAdV 2

The HAdV 2 E1A-12S construct is inserted in a pETG-20A EMBL plasmid [23] that was prepared by Gateway® cloning technology (Invitrogen). The HAdV 2 E1A-12S is a native 243 residues protein and was synthesized (GeneArt® Invitrogen) already inside the entry clone pENTR^{TM}/TEV/D-TOPO, containing TEV cleavage site and the E1A-12S gene of interest flanked by *att*L sites. The plasmid was modified to avoid the usual four residues N-terminal sequence (GSFT) in the final construct upon TEV cleavage and kanamycin resistance was included for further isolation. The cassette containing TEV+E1A-12S was transferred into destination vector pETG-20A (containing *att*R sites and ampicillin resitance) through the Gateway LR® reaction, to generate the expression clone of HAdV 2 E1A-12S protein fused with thioredoxin A and a His tag at the N terminus. Using this approach the final construct after TEV cleavage yields the native 243 residues E1A-12S protein.

The plasmid pETG-20A/HAdV 2 E1A-12S plasmid was transformed into *E. coli* BL21(DE3) (Stratagene) and a single colony was selected to prepare a pre-inoculum in LB with ampicillin, grown overnight at 37 °C, 180 rpm. The unlabeled, single labeled ^{15}N, and double labeled ^{13}C, ^{15}N cultures were inoculated with 10 ml pre-inoculum into 1 l minimal medium (48.5 mM Na_2HPO_4, 22.0 mM KH_2PO_4, 8.5 mM NaCl, 0.2 mM $CaCl_2$, 2.0 mM $MgSO_4$, 10.0 μM $ZnCl_2$, 1 mg/l each of biotin and thiamin, 7.5 mM $(NH_4)_2SO_4$ / $(^{15}NH_4)_2SO_4$ and 11.1 mM glucose / $^{13}C_6$-glucose) at 37 °C with constant agitation at 160 rpm. When optical density at 600 nm reached 0.6∼0.8, cells were induced for 5 h with IPTG (250 μM in case of unlabeled and single labeled ^{15}N cultures at 30 °C, and 100 μM in case of double labeled ^{13}C,^{15}N cultures at 37 °C). The cells were harvested by centrifugation at 8000g for 15 min and the cell pellet stored at -20 °C.

All buffers used in the E1A-12S protein purification were degassed and purification steps were taken inside an anaerobic chamber using refrigerated systems to keeps the

temperature lower then 10 °C. Frozen cells were thawed and suspended in 25 ml equilibration buffer A (50 mM Hepes pH 8.0, 500 mM NaCl, 5 mM imidazole, 2 mM DTT, 10 μM $ZnCl_2$ and a pill of protease inhibitors (Roche)). Cells were disrupted by sonication on ice (at 40 % sonication power) with cycles of 3 s with 10 s delay pulses for 15 min. Lysate cells were centrifuged at 30000 rpm for 30 min at 4 °C and the supernatant was loaded on a 5 ml Ni^{2+} HiTrap IMAC column (Amersham biosciences Europe) pre-equilibrated with buffer A. Column was washed using 10 times column volumes of buffer A with 50 mM imidazole and the E1A-12S protein was eluted using buffer A with 250 mM imidazole and concentrated till 2.5 ml. The eluted E1A-12S buffer was exchanged into 50 mM Tris pH 8.0, 200 mM NaCl, 5 mM imidazole, 10 mM DTT using a PD10 column (Amersham Biosciences) and 750 μl TEV (0.1 mg·ml^{-1}) added for overnight digestion at room temperature inside the anaerobic chamber. After digestion, the E1A-12S was purified using 5 ml Ni^{2+} HiTrap IMAC column pre-equilibrated with buffer B (10 mM HEPES pH 8.0, 100 mM NaCl, 5 mM imidazol, 2 mM DTT). The thioredoxin and histidine tags remained attached to the IMAC column and the digested E1A-12S was eluted in the flow-through. The E1A-12S was collected, warmed at 80 °C for 20 minutes and centrifuged at 4000 g for 15 minutes. The supernatant was filtrated, concentrated till 5 ml and loaded on a HiLoad 16/100 Superdex 75 (Amersham biosciences Europe) size-exclusion chromatography column previously equilibrated with 10 mM HEPES pH 7.5, 50 mM KCl, 10 mM DTT using a flow-rate of 1.0 ml·min^{-1}. The collected purified samples were checked by SDS-PAGE and mass-spectrometry to confirm the molecular mass and purity E1A-12S protein. The E1A-12S sample concentrations for NMR were 500 μM, estimated using the molar extinction coefficient (1490 M^{-1}cm^{-1}) calculated from ExPASy ProtParam tool [http://web.expasy.org/protparam/]. [24] The concentration was also confirmed by refractometer using a Rudolph Research Analytical J357 automatic refractometer.

Cloning, expression and purification of 13S isoform of E1A protein from HAdV 2/5

The HAdV 2/5 E1A-13S construct inserted in a pET42 (Novagen) plasmid was generously provided by Peter Pelka and Joe S. Mymryk, University of Western Ontario, London, Ontario, Canada. The plasmid construct was amplified and the DNA screened showing the sequence of the E1A-13S belonging to HAdV 2/5 hybrid. The final construct is basically the HAdV 2 E1A-13S protein sequence where SER 276 is mutated to PRO as in the HAdV 5 E1A-13S protein sequence and GLU 148 is mutated to GLY. The E1A-13S is 289 residues, which in this construct contains an additional C-terminus 8× histidine tail

with a linker (LEHHHHHHHH) resulting in a final construct with 299 residues.

The plasmid pET42/E1A-13S was transformed into *E. coli* BL21(DE3)pLyS (Stratagene) and a single colony was selected to prepare a pre-inoculum in LB with ampicillin and chloramphenicol, grown overnight at 37 °C, 180 rpm. The unlabeled, single labeled ^{15}N, and double labeled ^{13}C, ^{15}N cultures were inoculated with 10 ml pre-inoculum into 1 l minimal medium (48.5 mM Na_2HPO_4, 22.0 mM KH_2PO_4, 8.5 mM NaCl, 0.2 mM $CaCl_2$, 2.0 mM $MgSO_4$, 10.0 μM $ZnCl_2$, 1 mg/l each of biotin and thiamin, 7.5 mM $(NH_4)_2SO_4$ / $(^{15}NH_4)_2SO_4$ and 11.1 mM glucose / $^{13}C_6$-glucose) at 37 °C with constant agitation at 160 rpm. When optical density at 600 nm reached 0.6∼0.8, cells were induced with IPTG (500 μM in case of unlabeled, single labeled ^{15}N cultures and 100 μM in case of double labeled ^{13}C, ^{15}N) for additional 8 h at 25 °C. The cells were harvested by centrifugation at 8000 g for 15 min and the cell pellet stored at -20 °C.

All buffers used in the E1A-13S protein purification were degassed and purification steps were taken inside an anaerobic chamber using refrigerated systems to keeps the temperature lower than 10 °C. Frozen cells were thawed and suspended 25 ml equilibration buffer A (50 mM Hepes pH 8.0, 500 mM NaCl, 5 mM imidazole, 2 mM DTT, 10 μM $ZnCl_2$ and a pill of protease inhibitors (Roche)). Cells were disrupted by sonication on ice (at 40 % sonication power) with cycles of 1 s with 10 s delay pulses for 30 min. Lysate cells were centrifuged at 30000 rpm for 30 min at 4 °C and the supernatant was loaded on a 5 ml Ni^{2+} HiTrap IMAC column (Amersham biosciences Europe) pre-equilibrated with buffer A. Column was washed using 10 times column volumes of buffer A with 50 mM imidazole and the E1A-13S protein was eluted using buffer A with 250 mM imidazole. The eluted sample was filtrated, concentrated till 5 ml and loaded on a HiLoad 16/100 Superdex 75 (Amersham biosciences Europe) size-exclusion chromatography column previously equilibrated with 10 mM HEPES pH 7.5, 150 mM KCl, 10 mM DTT using a flow-rate of 1.0 ml·min^{-1}. The collected purified samples were checked by SDS-PAGE and mass-spectrometry to confirm the molecular mass and purity E1A-13S protein.

The E1A-13S sample concentrations for NMR were 250 μM, estimated using the molar extinction coefficient (7450 M^{-1}cm^{-1}) calculated from ExPASy ProtParam tool [http://web.expasy.org/protparam/]. [24] The concentration was also confirmed by refractometer using a Rudolph Research Analytical J357 automatic refractometer.

NMR Experiments

The ^{13}C detected NMR experiments for sequence specific assignment were acquired at 278 K on the 16.44 T Bruker AVANCE 700 NMR spectrometer, operating at 700.06 MHz

for ^{1}H, equipped with a cryogenically cooled probe optimized for ^{13}C direct detection. The ^{1}H detected NMR experiments for sequence specific assignment were acquired at 278 K on the 21.1 T Bruker AVANCE 900 NMR spectrometer, operating at 898.45 MHz for ^{1}H, equipped with cryogenically cooled TCI probe. A datasets of ^{13}C detected (CON-IPAP [25], 4D HCBCACON-IPAP [26], 4D HCBCANCO-IPAP [26], 4D HCACONCACON-IPAP [27], 4D HNCONCACON-IPAP [27]) and ^{1}H detected NMR experiments (1H-15N BEST-TROSY [28,29], BEST-TROSY HNCO [30–32], BEST-TROSY HNCACO [30–32], BEST-TROSY HNCOCACB [30–32], BEST-TROSY HNCACB [30–32], BEST-TROSY HN(CA)NNH [30, 32–34] and TROSY HN(COCA)NNH [35, 36]) were necessary for sequence specific assignment of H^{N}, N, C', C^{α}, C^{β} resonances. To characterize the structural and dynamic features of HAdV E1A, ^{15}N relaxation experiments (R_1, R_2 and ^{1}H-^{15}N NOEs) on a 16.44 T Bruker AVANCE 700 spectrometer equipped with a cryogenically cooled probe. All experimental parameters used to acquire NMR experiments described above are reported in more detail in the Supplementary Material (Table S1, S2, S3).

Data were processed with TopSpin 2.0 and were analyzed with program ccpNMR. [37] The secondary structure propensity from chemical shifts was determined using the ncSPC web program [www. http://nmr.chem.rug.nl/ncSPC/]. [38] The ^{15}N relaxation rates (R_1 and R_2) were determined by fitting the cross-peak intensity measured as a function of variable delay, to single-exponential decay. ^{1}H-^{15}N NOE values were obtained as a ratio between peak intensity in spectra recorded with and without ^{1}H saturation.

Circular Dichroism

CD experiments were performed on a Jasco J-810 spectropolarimeter (Easton, MD), equipped with a PTC-423S Peltier temperature control unit, using a cuvette with 1 mm path length. Hellmanex II reagent (Hellma GmbH & Co. KG) in a concentration of 2 % in water was used to clean the cuvette one day before acquiring the CD spectra. Spectra of 15.84 μM E1A-12S (concentration determined by the J357 Automatic Refractometer - Rudolph Research) and 15.85 μM E1A-13S (concentration determined by the Cary 50 Scan UV-Visible Spectrophotometer) acquired in a 180-250 nm wavelength in 10 mM Tris-Mes, 30 mM Na_2SO_4 without 10 μM Zinc Acetate (E1A-12S) and with 10 μM Zinc Acetate (E1A-13S) at three different pH values (6.0, 7.65, 9.0), were recorded at 4 °C. All the data sets were collected using Spectra Manager software, 30 numbers of scans and the CD data in millidegrees (Θ). The buffers were filtered through a 0.2 μm filter and the protein samples were put into the cuvettes inside the glove box. BSA was used as a control of the experiments. It was prepared at the same buffers used for the protein samples.

Results and Discussions

Amino acid sequence alignment of E1A-12S and E1A-13S nicely depicts the presence of three conserved regions (CR1,2,4) common to the two variants, while a thirds conserved region (CR3) is present only in the longer variant, as schematically indicated in Figure 1. Small segments of non-conserved regions are also present, [39] which are also involved in important protein-protein interactions with over 50 cellular factors through SLiMs. [11,21] The SLiMs identified in the conserved regions are indicated in black in Figure 1. The high resolution structural and dynamic characterization of E1A represents an important step forward in the understanding of the molecular basis of its functional properties by providing experimental data to be compared with predictions and with a wealth of complementary information available on the protein through different molecular biology and biophysical techniques. [21]

The primary sequence of E1A-12S (243 residues) and E1A-13S (289 residues) can be used to predict through well-established bioinformatics tools the tendency of the different parts of the polypeptide chain to adopt more ordered or disordered conformations and which are likely interaction sites. As an example, predictions performed with PONDR-FIT [40] and Anchor [41] are reported in the Supplementary Material (Figure S1). Both proteins are expected to be characterized by significant structural heterogeneity, with largely disordered parts as well as small regions prone to adopt transient structures. The E1A-13S isoform contains in addition the ordered CR3 region, which can be clearly identified in the PONDR-FIT graph (Figure S1).

Circular dichroism (CD) experiments were used to obtain hints on the secondary structure composition of the two isoforms. The CD spectra of both E1A-12S and E1A-13S are shown in Figure 2. The CD spectra of E1A-12S (Figure 2A) at three different pH values (6.0, 7.6 and 9.0) at 278 K display a large and negative signal at 190 nm and negligible ellipticities at 208 nm and 222 nm, consistent with the completely unfolded polypeptide. [42] At pH 6.0 also the E1A-13S spectrum shows large and negative signal at 190 nm and negligible ellipticities at 208 nm and 222 nm, but at higher pH values a positive signal is observed at interval from 188 nm to 200 nm, which is indicative for the presence of secondary structure elements (Figure 2). These findings are in line with the presence of the structured CR3 region in the E1A-13S form. It also suggests that the fold of the CR3 region is stabilized by the tighter binding of the zinc ion by two CxxC motifs at pHs above 8, which results in lower tendency to aggregate.

The 2D ^{1}H-^{15}N correlation experiments acquired on E1A-12S and E1A-13S are shown

in Figure 3. The low chemical shift dispersion of the cross peaks observed confirms that both isoforms of E1A are highly disordered and flexible, as inferred from bioinformatic predictions (Figure S1). Thus, the sequence specific assignment of the resonances represents a challenging task, both due to the extensive overlap observed as well as due to structurally heterogeneous nature of E1A.

The first obvious difference in behavior of the two studied isoforms is their tendency to aggregate, which impose different NMR strategies. The tendency of the E1A-13S isoform to from large aggregates upon increasing concentration poses limits in terms of protein concentrations that can be used for NMR experiments, restricting the choice among the available assignment strategies to those including the most sensitive ^{1}H detected NMR experiments. The E1A-12S isoform doesnt show tendency to aggregate, hence higher concentration of the protein could be achieved, which allowed us to use ^{13}C direct detected experiments for sequence specific assignment.

The analysis of experiments exploiting the higher chemical shift dispersion provided by the carbonyl dimension, enables us to count the number of cross peaks that can be detected. In the case of the E1A-12S, 240 isolated cross peaks could be identified in the CON experiment, which corresponds what is expected on the basis of the primary sequence. (Figure S1 a) Thus, a series of 4D ^{13}C-detected experiments based on carbonyl detection were selected for the sequence specific assignment (4D HCBCACON-IPAP, 4D HCBCANCO-IPAP, 4D HCACONCACON -IPAP, 4D HNCONCACON-IPAP). In the case of the E1A-13S protein only 120 cross peaks (out of 288 expected) were counted in the CON experiment (Figure S1 b), while 200 signals were observed in the HNCO experiment, which is significantly less than what expected on the basis of the primary sequence but still represents more than 80 % of the number of residues that can be studied by ^{1}H detected experiments. Thus, the strategy based on $^{1}H^{N}$-detected experiments was chosen for sequence specific assignment of the E1A-13S isoform. In addition to the widely used 3D experiments (HNCO/HNCACO, CBCANH, CBCACONH) also two additional experiments, which provide sequential correlations exploiting backbone nitrogen atoms and carbonyl carbons and which are known to provide valuable information to resolve ambiguities in particular when investigating highly disordered protein segments (HNCANNH, HNCOCANNH), were selected. The BEST-TROSY versions of the experiments were used. The complete set of experiment recorded on the two E1A variants and the corresponding parameters are reported as Supplementary Material (Table S1). The results of the sequence specific assignment were deposited to BMRB (to be deposited).

The analysis of the spectra allowed us to achieve essentially complete sequence specific

assignment of the polypeptide chain comprising CR1 and CR2 as well as the final module comprising CR4 for the both isoforms. The assigned residues of the E1A-12S include 97 % of the residues involved in interactions with protein partners, as identified by ANCHOR, and all of the residues which are part of the various SLiMs predicted in the primary sequence. In the case of the E1A-13S protein, acquired experiments are blind towards the structured CR3 region and the neighboring residues in its vicinity supporting the assumption of CR3 be responsible for aggregates formation.

The sequence specific assignment provides the first information to characterize the structural and dynamic properties of the protein through the analysis of backbone heteronuclear chemical shifts and of those of C^{β} carbon nuclei. Indeed by comparing experimental chemical shifts with those predicted assuming a completely random coil conformation it is possible to identify if secondary structural elements are partially populated. The resulting CSI values for C^{α}, C^{β}, C' and N as well as the SSP scores are reported in Figure 4.

Further observables providing useful information to characterize the structural and dynamic properties of the protein are ^{15}N relaxation data, in particular ^{15}N R_1, ^{15}N R_2 relaxation rates and ^{1}H-^{15}N heteronuclear NOEs, reported in Figure 5. These show average values of 1.6 Hz for R_1 and of 6.0 Hz for R_2 for E1A-12S as well as for E1A-13S (regions showing significantly higher or lower ^{15}N R_2 were excluded from average calculation), which indicate that both isoforms are largely disordered. In addition the ^{1}H-^{15}N NOE values, all significantly lower than 0.8, also confirm that the proteins is largely disordered. However, inspection of all the data does indicate that the proteins are characterized by a significant structural and dynamic heterogeneity, with several regions characterized by a propensity to populate more structured and/or less flexible conformations.

The first evidence of a region characterized by a more structured conformation comes from the analysis of chemical shifts. Indeed both CSI as well as SSP do indicate that in the initial, N-terminal part of the E1A protein, amino acids 13-23 have a significant α-helical secondary structural propensity. This is confirmed by reduced local mobility in this region as identified through ^{15}N relaxation measurements, with R_2 values higher than the average (11-21) as well as with larger ^{1}H-^{15}N NOEs (11-20). The further inspection of CSI shows that also region from residue number 132 to 140 has slight propensity to adopt extended conformation. The rest of the polypeptide chain instead shows smaller CSI and SSP values.

However, inspection of the ^{15}N relaxation data does show several regions characterized by R_2 and NOE values larger than average, which in general are indicative of less flexible

parts of the polypeptide chain (120-126 for both isoforms; regions surrounding CR3 in E1A-13S; regions 210-224 and 258-270 of E1A-13S, resp. 158-173 and 212-224 in E1A-12S isoform). More interestingly, the ^{15}N relaxation data reveal the differences in the behavior of the E1A-12S and E1A-13S isoform. The most striking difference is significantly higher R_2 values in the regions of E1A-13S from amino acids 120-137, the region containing several SLiMs, as for example the one identified as the interaction site for the retinoblastoma binding protein (pRb), and also different profiles of R_2 values in CR4 region. Overall, identified regions with the different local dynamics between the two forms are mainly in the vicinity of the reported SLiMs, which could have an effect on their binding to the cellular targets.

Till now, only several short fragments of the E1A were used for high-resolution interaction studies. The interaction of the initial part of the polypeptide chain with the TAZ2 module of the CREB binding protein (CBP) has been investigated by analyzing several short constructs and revealed that different regions are involved in binding to the partner with different properties. [43] These regions include the first part of the polypeptide where a high α-helical propensity has been identified (13-23) as well as two other regions comprising residues 59-65 and 72-75 that adopt an alpha helical conformation upon binding to the partner. The LxCxE short linear motif (122-126), located in CR2, is known to interact with the retinoblastoma binding protein (pRb), and promote its degradation. [11] The interaction of a short peptide mimicking this part of the polypeptide chain from the homologous protein E7 from the human papilloma virus with pRb has been characterized and showed that the LxCxE motif adopts an elongated conformation upon binding to pRb. [44] Interestingly the investigation of full length E7 showed a mild propensity of this part of the backbone as well as of neighboring ones, to assume an elongated conformation.

In the present case, we observed similar behavior, but ^{15}N relaxation data suggests higher flexibility of the regions in the E1A-12S compared to the E1A-13S protein, which could have an effect on affinity of the two isoforms. Thus it can be an indication of the cross-talk between structured CR3 and other regions that could probably also partially contribute to the slightly different biological activity of the E1A-12S and E1A-13S isoform at the different stages of the infection (reviewed by Chang et al. [45]).

Conclusions

In summary, we have shown that with an integrated NMR approach including ^{1}H and ^{13}C detected multidimentional triple resonance experiment, it is possible to characterize

at high resolution the two major forms of E1A from HAdV. Thanks to vast amount of data available on the functional properties of this protein, of its interactions with many partners involved in key processes in the cell, we trust the present results will provide the basis for the high resolution characterization of the interactions in which this protein is known to be involved.

Acknowledgments

This work has been supported in part by the Joint Research Activity and Access to Research Infrastructures (BioNMR, contract 261863) and by the Marie Curie ITN programs (IDPbyNMR, contract 264257) in the EC 7th Framework.

Supplementary Material

Acquisition parameters of the NMR experiments; comparison of 2D HN and CON maps; bioinformatics prediction on the extent of disorder; sequence specific assignment of both isoforms.

References

[1] Wagner G, Hyberts S, Peng JW. Study of protein dynamics by NMR. In: *NMR of proteins*, edited by Clore GM, Gronenborn AM, pp. 220–257. Boca Raton: CRC Press Inc. 1993;.

[2] Wagner G. NMR investigations of protein structure. *Progress in Nuclear Magnetic Resonance Spectroscopy*. 1990;22(2):101–139.

[3] Mittermaier A, Kay LE. New tools provide new insights in NMR studies of protein dynamics. *Science*. 2006;312(5771):224–228.

[4] Wright PE, Dyson HJ. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *Journal of Molecular Biology*. 1999;293(2):321–331.

[5] He B, Wang K, Liu Y, Xue B, Uversky VN, Dunker AK. Predicting intrinsic disorder in proteins: an overview. *Cell Res*. 2009;19(8):929–949.

[6] Tompa P. Multisteric regulation by structural disorder in modular signaling proteins: an extension of the concept of allostery. *Chem Rev*. 2014;114(13):6715–6732.

[7] Uversky VN, Oldfield CJ, Dunker AK. Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J Mol Recognit*. 2005;18(5):343–384.

[8] Fuxreiter M, Simon I, Friedrich P, Tompa P. Preformed Structural Elements Feature in Partner Recognition by Intrinsically Unstructured Proteins. *Journal of Molecular Biology*. 2004;338(5):1015–1026.

[9] Uversky VN, Oldfield CJ, Dunker AK. Intrinsically Disordered Proteins in Human Diseases: Introducing the D 2Concept. *Annu Rev Biophys*. 2008;37(1):215–246.

[10] Gould CM, Diella F, Via A, Puntervoll P, Gemünd C, Chabanis-Davidson S, Michael S, Sayadi A, Bryne JC, Chica C, Seiler M, Davey NE, Haslam N, Weatheritt RJ, Budd A, Hughes T, Pas J, Rychlewski L, Travé G, Aasland R, Helmer-Citterich M, Linding R, Gibson TJ. ELM: the status of the 2010 eukaryotic linear motif resource. *Nucleic Acids Research*. 2010;38:D167–D180.

[11] Davey NE, Travé G, Gibson TJ. How viruses hijack cell regulation. *Trends in Biochemical Sciences*. 2011;36(3):159–169.

[12] Wells M, Tidow H, Rutherford TJ, Markwick P, Jensen MR, Mylonas E, Svergun DI, Blackledge M, Fersht AR. Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. *Proc Natl Acad Sci USA*. 2008; 105(15):5762–5767.

[13] Romero PR, Zaidi S, Fang YY, Uversky VN, Radivojac P, Oldfield CJ, Cortese MS, Sickmeier M, LeGall T, Obradovic Z, Dunker AK. Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. *Proc Natl Acad Sci USA*. 2006;103(22):8390–8395.

[14] Gallimore PH. Tumour production in immunosuppressed rats with cells transformed in vitro by adenovirus type 2. *J Gen Virol*. 1972;16(1):99–102.

[15] Stewart PL, Burnett RM, Cyrklaff M, Fuller SD. Image reconstruction reveals the complex molecular organization of adenovirus. *Cell*. 1991;67(1):145–154.

[16] Mackey JK, Rigden PM, Green M. Do highly oncogenic group A human adenoviruses cause human cancer? Analysis of human tumors for adenovirus 12 transforming DNA sequences. *Proc Natl Acad Sci USA*. 1976;73(12):4657–4661.

[17] Mackey JK, Wold WS, Rigden P, Green M. Transforming region of group A, B, and C adenoviruses: DNA homology studies with twenty-nine human adenovirus serotypes. *Journal of Virology*. 1979;29(3):1056–1064.

[18] Jones N, Shenk T. An adenovirus type 5 early gene function regulates expression of other early viral genes. *Proc Natl Acad Sci USA*. 1979;76(8):3665–3669.

[19] Nevins JR, Ginsberg HS, Blanchard JM, Wilson MC, Darnell JE. Regulation of the primary expression of the early adenovirus transcription units. *Journal of Virology*. 1979;32(3):727–733.

[20] Hearing P, Shenk T. The adenovirus type 5 E1A transcriptional control region contains a duplicated enhancer element. *Cell*. 1983;33(3):695–703.

[21] Pelka P, Ablack JNG, Fonseca GJ, Yousef AF, Mymryk JS. Intrinsic Structural Disorder in Adenovirus E1A: a Viral Molecular Hub Linking Multiple Diverse Processes. *Journal of Virology*. 2008;82(15):7252–7263.

[22] Chinnadurai G. Opposing oncogenic activities of small DNA tumor virus transforming proteins. *Trends Microbiol*. 2011;19(4):174–183.

[23] Alzari PM, Berglund H, Berrow NS, Blagova E, Busso D, Cambillau C, Campanacci V, Christodoulou E, Eiler S, Fogg MJ, Folkers G, Geerlof A, Hart D, Haouz A, Herman MD, Macieira S, Nordlund P, Perrakis A, Quevillon-Cheruel S, Tarandeau F, van Tilbeurgh H, Unger T, Luna-Vargas MPA, Velarde M, Willmanns M, Owens RJ. Implementation of semi-automated cloning and prokaryotic expression screening: the impact of SPINE. *Acta Crystallogr D Biol Crystallogr*. 2006;62(Pt 10):1103–1113.

[24] Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A. Protein identification and analysis tools on ExPASy server. In: *The Proteomics Protocols Handbook*, edited by Walker JM, pp. 571–607. Humana Pr Inc. 2005;.

[25] Bermel W, Bertini I, Duma L, Felli IC, Emsley L, Pierattelli R, Vasos PR. Complete Assignment of Heteronuclear Protein Resonances by Protonless NMR Spectroscopy. *Angew Chem Int Ed*. 2005;44(20):3089–3092.

[26] Bermel W, Bertini I, Felli IC, Gonnelli L, Koźmiński W, Piai A, Pierattelli R, Stanek J. Speeding up sequence specific assignment of IDPs. *J Biomol NMR*. 2012;53(4):293–301.

[27] Bermel W, Felli IC, Gonnelli L, Koźmiński W, Piai A, Pierattelli R, Zawadzka-Kazimierczuk A. High-dimensionality C-13 direct-detected NMR experiments for the automatic assignment of intrinsically disordered proteins. *J Biomol NMR*. 2013; 57(4):353–361.

[28] Favier A, Brutscher B. Recovering lost magnetization: polarization enhancement in biomolecular NMR. *J Biomol NMR*. 2011;49(1):9–15.

[29] Lescop E, Kern T, Brutscher B. Guidelines for the use of band-selective radiofrequency pulses in hetero-nuclear NMR: example of longitudinal-relaxation-enhanced BEST-type 1H-15N correlation experiments. *J Magn Reson*. 2010;203(1):190–198.

[30] Lescop E, Schanda P, Brutscher B. A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *Journal of Magnetic Resonance*. 2007; 187(1):163–169.

[31] Salzmann MM, Pervushin K, Wider GG, Senn HH, Wüthrich K. TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. *Proc Natl Acad Sci USA*. 1998;95(23):13585–13590.

[32] Solyom Z, Schwarten M, Geist L, Konrat R, Willbold D, Brutscher B. BEST-TROSY experiments for time-efficient sequential resonance assignment of large disordered proteins. *J Biomol NMR*. 2013;55(4):311–321.

[33] Schanda P, Van Melckebeke H, Brutscher B. Speeding up three-dimensional protein NMR experiments to a few minutes. *J Am Chem Soc*. 2006;128(28):9042–9043.

[34] Weisemann R, Rüterjans H, Bermel W. 3D Triple-resonance NMR techniques for the sequential assignment of NH and 15N resonances in 15N- and 13C-labelled proteins. *J Biomol NMR*. 1993;3(1):113–120.

[35] Bracken C, Palmer AG III, Cavanagh J. (H)N(COCA)NH and HN(COCA)NH experiments for 1H-15N backbone assignments in 13C/15N-labeled proteins. *J Biomol NMR*. 1997;9(1):94–100.

[36] Panchal SC, Bhavesh NS, Hosur RV. Improved 3D triple resonance experiments, HNN and HN(C)N, for HN and 15N sequential correlations in (13C, 15N) labeled proteins: application to unfolded proteins. *J Biomol NMR*. 2001;20(2):135–147.

[37] Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J, Laue ED. The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins*. 2005;59(4):687–696.

[38] Tamiola K, Mulder FAA. Using NMR chemical shifts to calculate the propensity for structural order and disorder in proteins. *Biochemical Society transactions*. 2012; 40(5):1014–1020.

[39] Kimelman D, Miller JS, Porter D, Roberts BE. E1a regions of the human adenoviruses and of the highly oncogenic simian adenovirus 7 are closely related. *Journal of Virology*. 1985;53(2):399–409.

[40] Xue B, Dunbrack RL, Williams RW, Dunker AK, Uversky VN. PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochim Biophys Acta*. 2010; 1804(4):996–1010.

[41] Dosztányi Z, Mészáros B, Simon I. ANCHOR: web server for predicting protein binding regions in disordered proteins. *J Gerontol*. 2009;25(20):2745–2746.

[42] Uversky VN. What does it mean to be natively unfolded? *Eur J Biochem*. 2002; 269(1):2–12.

[43] Ferreon JC, Martinez-Yamout MA, Dyson HJ, Wright PE. Structural basis for subversion of cellular control mechanisms by the adenoviral E1A oncoprotein. *Proc Natl Acad Sci USA*. 2009;106(32):13260–13265.

[44] Lee JO, Russo AA, Pavletich NP. Structure of the retinoblastoma tumour-suppressor pocket domain bound to a peptide from HPV E7. *Nature*. 1998;391(6670):859–865.

[45] Chang YW, Hung MC, Su JL. The Anti-Tumor Activity of E1A and its Implications in Cancer Therapy. *Arch Immunol Ther Exp (Warsz)*. 2014;62(3):195–204.

[46] Kjaergaard M, Brander S, Poulsen FM. Random coil chemical shift for intrinsically disordered proteins: effects of temperature and pH. *J Biomol NMR*. 2011;49(2):139–149.

Figures

Figure 1. Schematic representation of HAdV 2 E1A-12S, resp. HAdV 2/5 E1A-13S, protein amino acid sequence, and division into 3, resp. 4, conserved region (CR) with short linear motifs (SLiMs) indicated in black. [21]

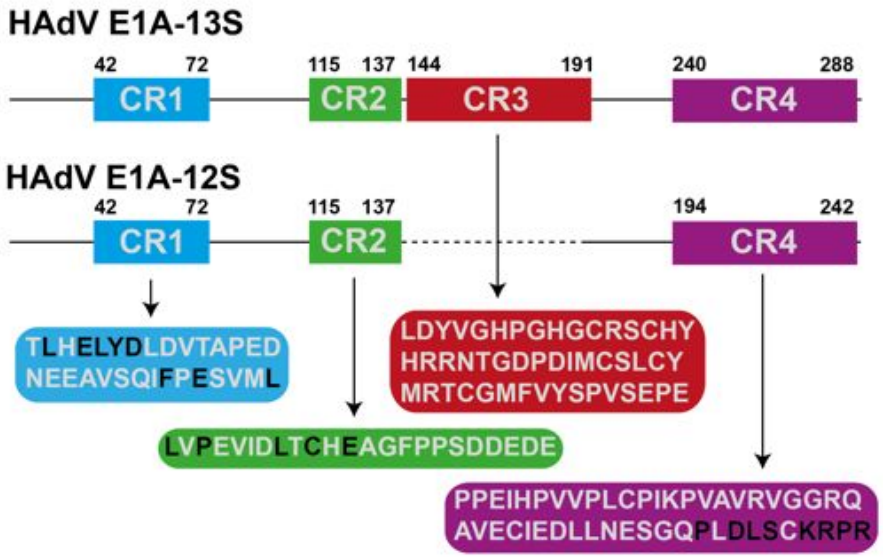


Figure 2. Far-UV CD spectra of the HAdV 2 E1A-12S (a) and HAdV 2/5 E1A-13S (b) recorded at pH 6.0 (light gray traces), 7.65 (dark gray traces) and 9.0 (black line). The data were collected at 4 °C, in 10 mM Tris-Mes, 30 mM Na_2SO_4 without 10 μM Zinc Acetate (E1A-12S) and with 10 μM Zinc Acetate (E1A-13S). Protein concentration was 15.84 μM for E1A-12S and 15.85 μM for E1A-13S.

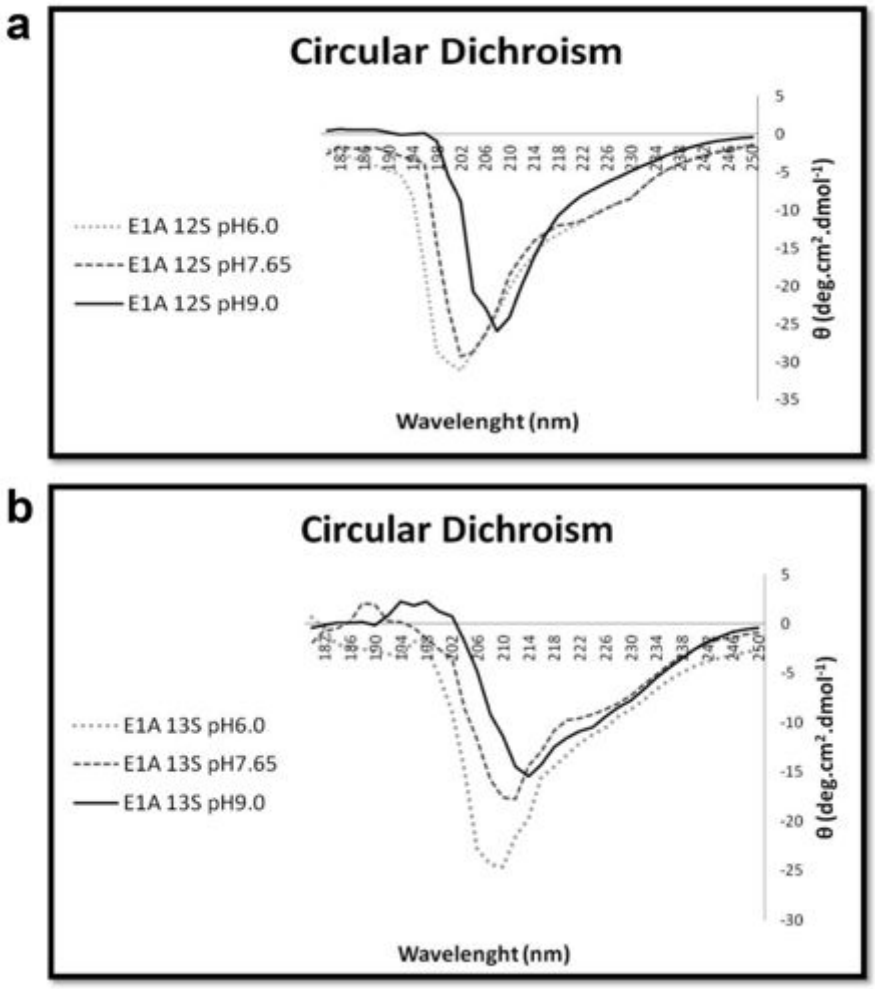


Figure 3. The 2D ^{1}H-^{15}N BEST-TROSY spectra recorded for HAdV 2 E1A-12S (a) and HAdV 2/5 E1A-13S (b). Experiments were acquired at 21.1 T NMR spectrometer at 278 K on 0.5 mM E1A-12S sample in 10 mM HEPES, pH 7.5, 50 mM KCl, and 10 mM DTT, and 0.2 mM E1A-13S sample in 10 mM HEPES, pH 7.5, 150 mM KCl, 10 μM $ZnCl_2$, and 10 mM DTT.

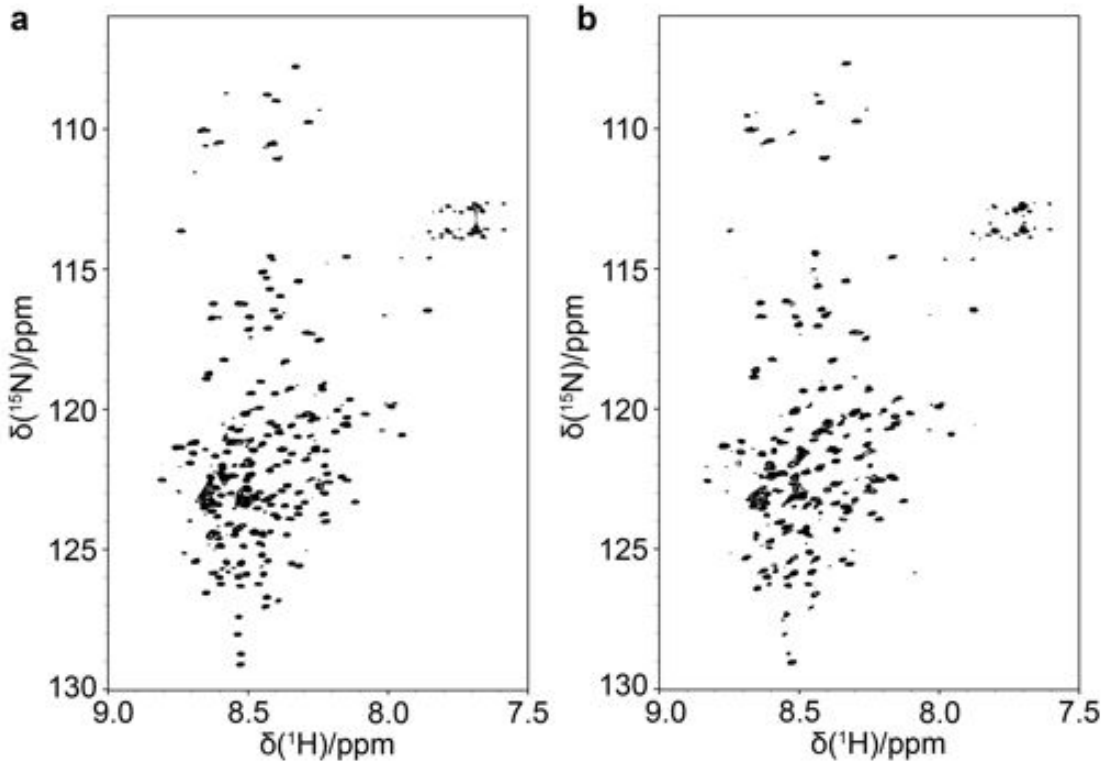
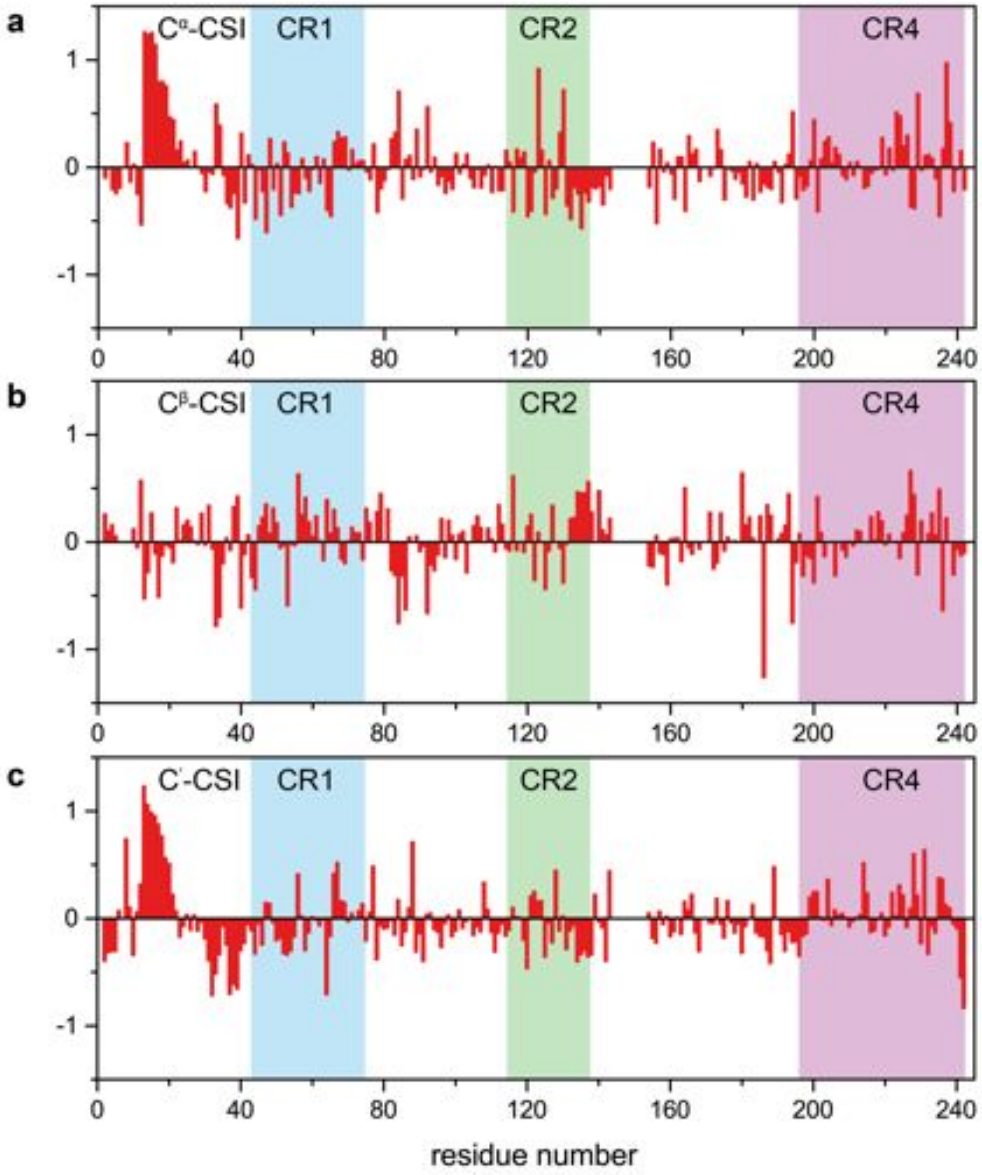
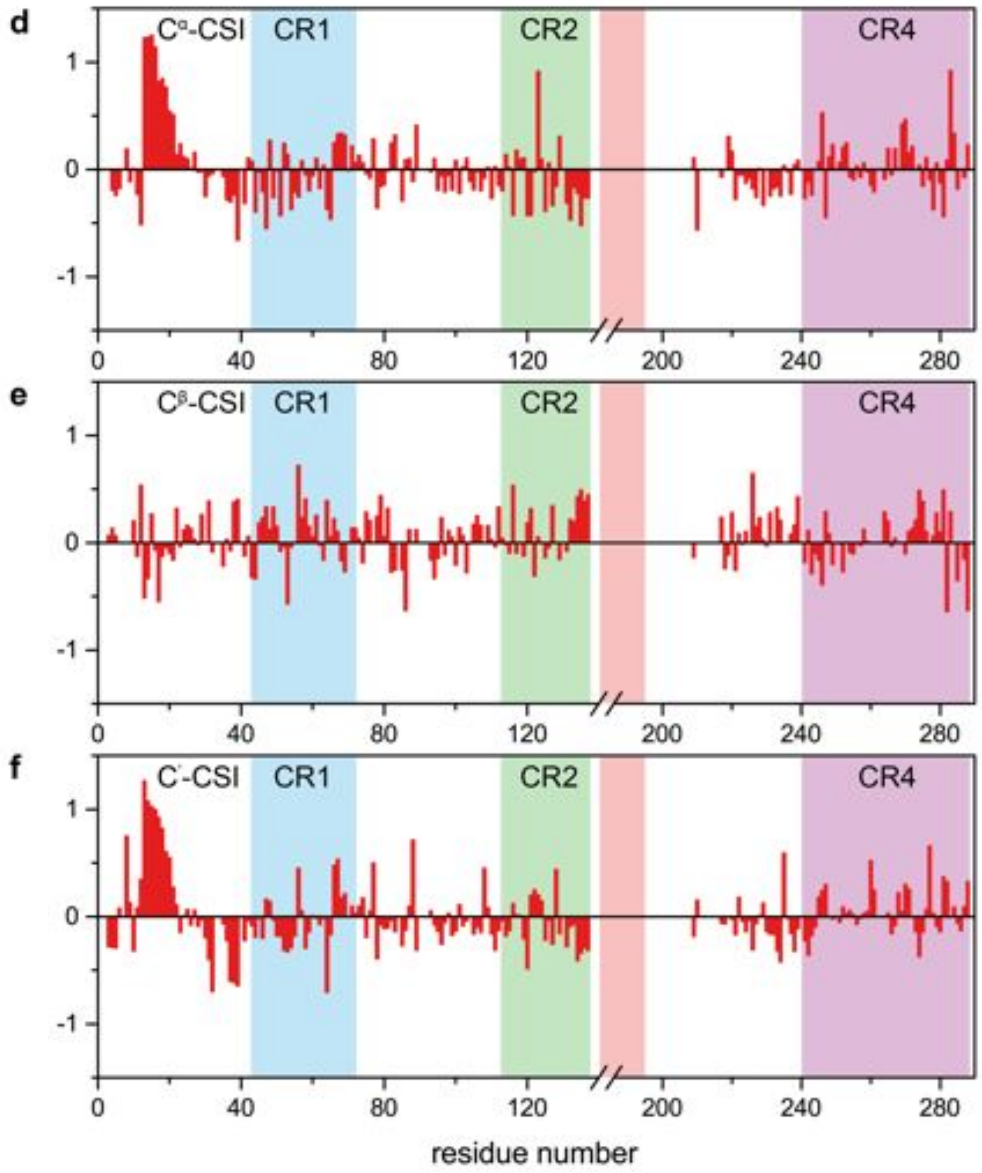


Figure 4. The chemical shift differences between the assigned resonances of HAdV 2 E1A-12S (a-c), resp. HAdV 2/5 E1A-13S (d-f), with respect to random coil values corrected for the sequence, pH and temperature [46] for , C^{α} (a, d), C^{β} (b, e))and C' (c, f) nuclei. The Secondary structure propensity (SSP) score [38] obtained by combining chemical shifts for the all assigned resonances of HAdV 2 E1A-12S (g) and . HAdV 2/5 E1A-13S (h)




d
Cα-CSI
CR1
CR2
CR4
1
0
-1
0
40
80
120
200
240
280
e
Cβ-CSI
CR1
CR2
CR4
1
0
-1
0
40
80
120
200
240
280
f
C'-CSI
CR1
CR2
CR4
1
0
-1
0
40
80
120
200
240
280
residue number

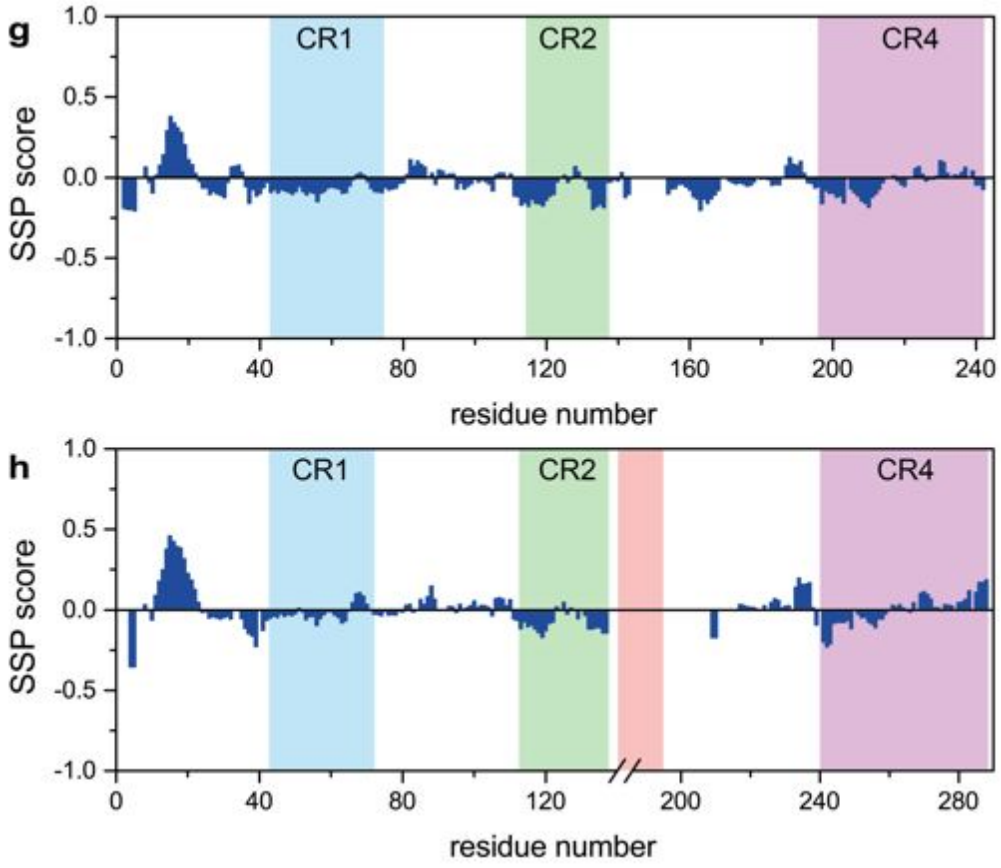
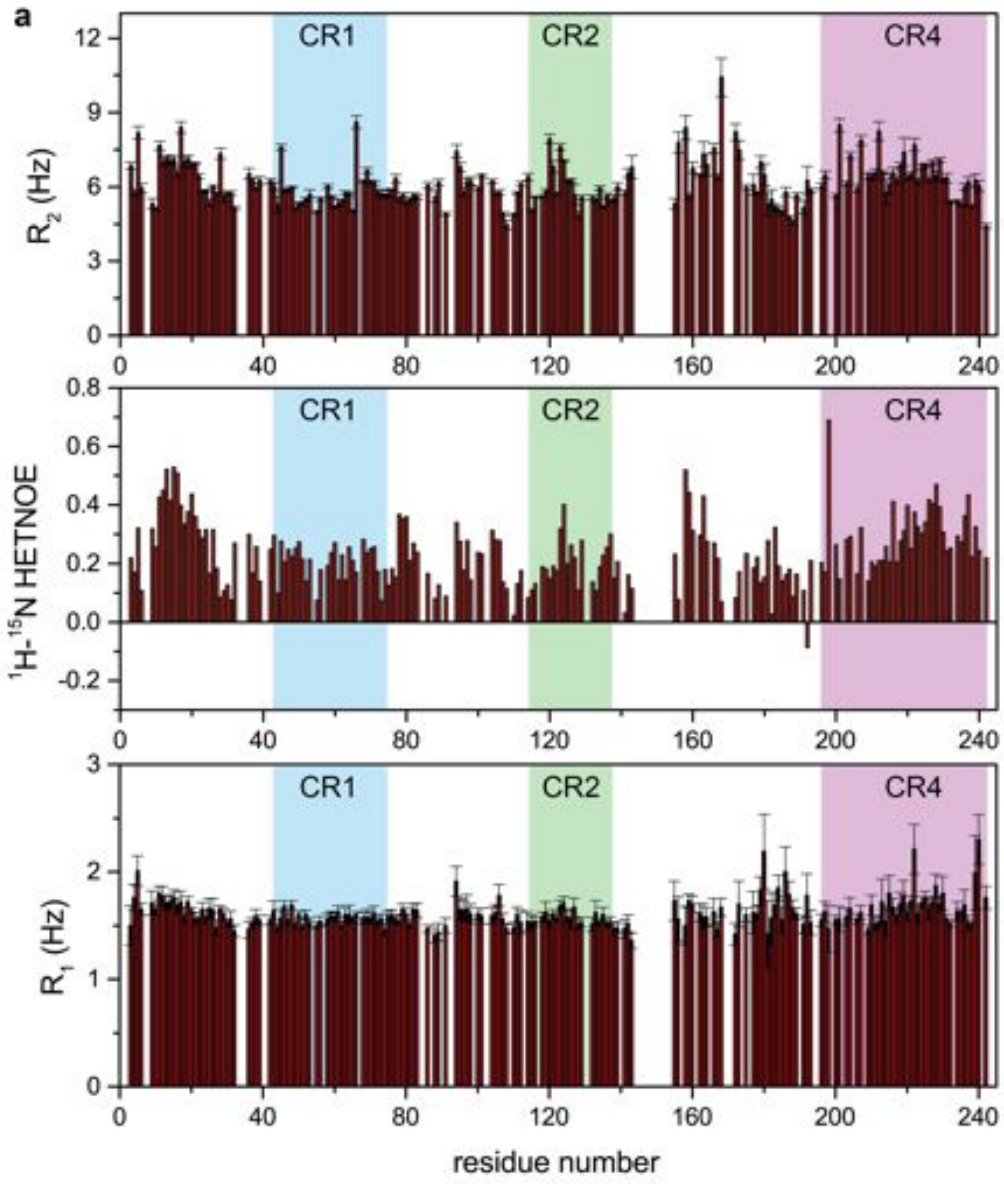
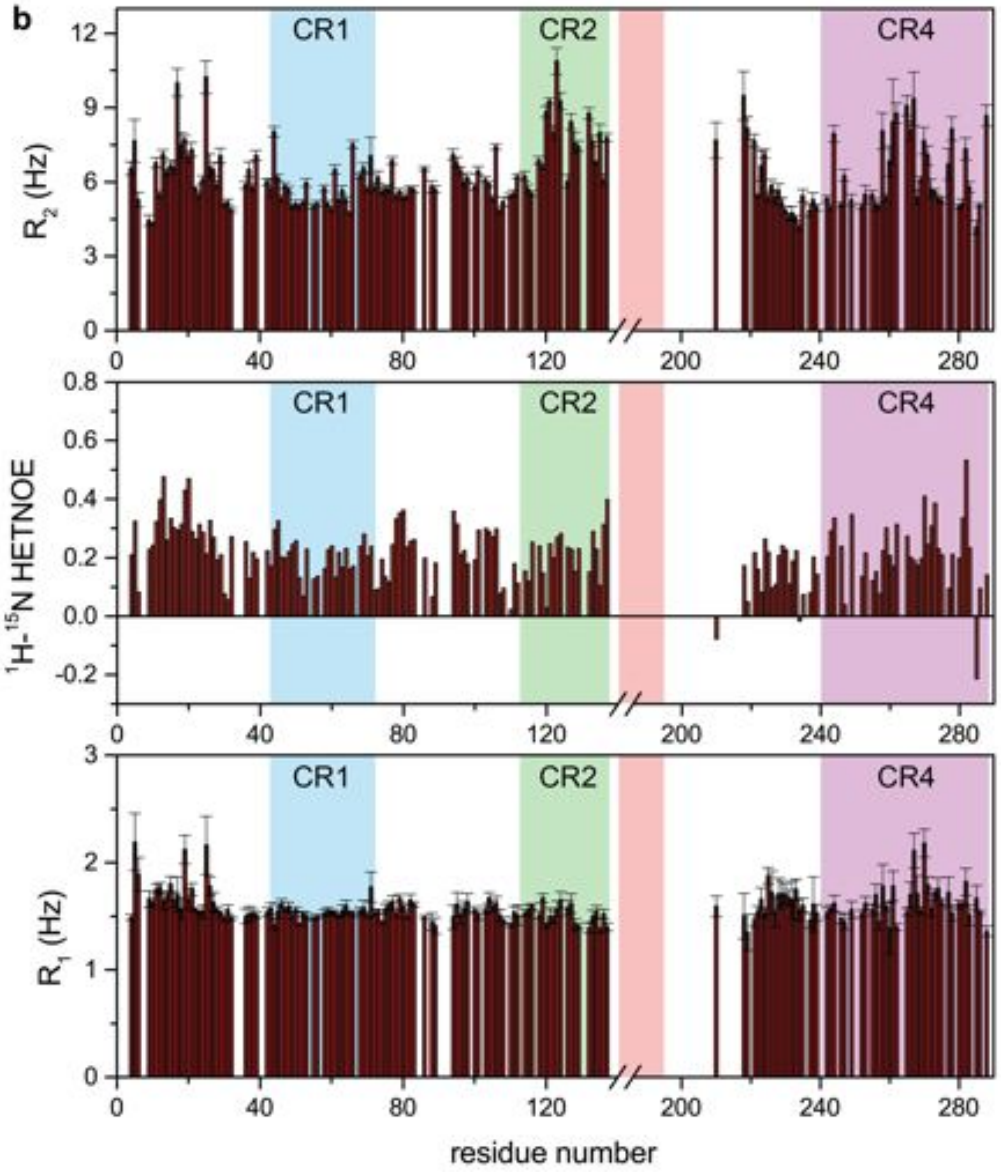

g
CR1
CR2
CR4
SSP score
1.0
0.5
0.0
-0.5
-1.0
0
40
80
120
160
200
240
residue number
h
CR1
CR2
CR4
SSP score
1.0
0.5
0.0
-0.5
-1.0
0
40
80
120
200
240
280
residue number

Figure 5. Relaxation data for backbone amide ^{15}N nuclei of HAdV 2 E1A-12S (a), resp. HAdV 2/5 E1A-13S (b), as function of the residue number. From top to bottom: ^{15}N longitudinal relaxation rates (R_1), ^{15}N transverse relaxation rates (R_2), and ^{1}H-^{15}N NOE values (HETNOE). Experiments were recorded on 16.44 T spectrometer at 278 K.




b
CR1
CR2
CR4
R2 (Hz)
1H-15N HETNOE
R1 (Hz)
residue number

The highly flexible and heterogeneous nature of HAdV E1A characterized at atomic resolution through NMR

Tomáš Hošek[a], Eduardo O. Calçada[a], Michele Salvi[a], Talita Pagani[a], Isabella C. Felli[a], Roberta Pierattelli[a]

[a] *CERM and Department of Chemistry Ugo Schiff, University of Florence, via Luigi Sacconi 6, Sesto Fiorentino, 50019, Italy*

SUPPLEMENTRY MATERIAL

Acquisition Parameters

The NMR experiments and acquisition parameters used for sequence specific assignment of H^N, N, C', C^α, C^β resonances and for further dynamic characterization of HAdV 2/5 E1A-13S and HAdV 2 E1A-12S at 278 K.

The ^{13}C detected experiments for sequence specific assignment were acquired with a Bruker AVANCE 700 spectrometer equipped with a cryogenically cooled probe optimized for ^{13}C direct detection (CON-IPAP [1], 4D HCBCACON-IPAP [2], 4D HCBCANCO-IPAP [2], 4D HCACONCACON-IPAP [3] and 4D HNCONCACON-IPAP [3]). The ^{13}C carrier was placed at 173 ppm, the ^{15}N carrier at 124.8 ppm and the ^{1}H carrier at 4.7 ppm. Band selective ^{13}C pulses were given at 173 and 39 ppm to excite or invert C' and C^{ali} spins respectively. The following band-selective pulses were used: 300 μs with newQ5 and time reversed newQ5 shapes for C' and C^{ali} excitation, 220 μs newQ3 shape for C'/C^{ali} inversion.

The ^{1}H detected experiments for sequence specific assignment were acquired with a Bruker AVANCE 900 spectrometer equipped with an inverse triple-resonance cryogenically cooled probehead (^{1}H-^{15}N BEST-TROSY [4, 5], BEST-TROSY HNCO [6–8], BEST-TROSY HNCACO [6–8], BEST-TROSY HNCOCACB [6–8], BEST-TROSY HNCACB [6–8], BEST-TROSY HN(CA)NNH [6,9,10] and TROSY HN(COCA)NNH [11,12]). The ^{1}H carrier was placed at 4.7 ppm, the ^{15}N carrier at 118.5 ppm and the ^{13}C carrier at 173 ppm for C' and at 39 ppm for C^{ali} . Band selective ^{1}H pulses were given at 8.9 ppm to excite or invert H^N spins. The following band-selective pulses were used: 1800 μs with Pc9 shape for H^N excitation, 1270 μs EBURP and time reversed EBURP shapes for H^N excitation, 1250 μs REBURP shape for H^N inversion. Band selective ^{13}C pulses were given at 173 and 39 ppm to excite or invert C' and C^{ali} spins respectively. The following band-selective pulses were used: 274 μs with G4 and time reversed G4 shapes for C' and C^{ali} excitation, 190 μs Q3 shape for C'/C^{ali} inversion.

The other experiments (^{15}N R_1, ^{15}N R_2 and ^{1}H-^{15}N NOEs) were acquired with a Bruker AVANCE 700 spectrometer equipped with an inverse triple-resonance cryogenically cooled probehead.

Table S1: Experiments used for HAdV 2/5 E1A-13S sequence specific assignment.

Experiments used for HAdV 2/5 E1A-13S sequence specific assignment	**Dimension of acquired data**			**Spectral width (ppm)**			n[a]	d[b]
	t_1	t_2	t_3	F_1	F_2	F_3		
^{13}C detected								
CON-IPAP[c]	512*(^{15}N)	1024(^{13}C)		38	30		48	2.2
HCBCACON-IPAP[c]	128*(^{13}C)	64(^{15}N)	1024(^{13}C)	60	30	49	32	1
^{1}H detected								
^{1}H-^{15}N HSQC	256(^{15}N)	1024(^{1}H)		23	12		4	1
^{1}H-^{15}N BEST-TROSY	512(^{15}N)	2048(^{1}H)		30	12		8	0.2
BEST-TROSY HNCO	96(^{13}C)	196(^{15}N)	2048(^{1}H)	8	24	15	8	0.2
BEST-TROSY HN(CA)CO	164(^{13}C)	196(^{15}N)	2048(^{1}H)	8	24	15	16	0.2
BEST-TROSY HN(CO)CACB	164(^{13}C)	196(^{15}N)	2048(^{1}H)	70	24	15	16	0.2
BEST-TROSY HNCACB	164(^{13}C)	196(^{15}N)	2048(^{1}H)	70	24	15	16	0.2
BEST-TROSY HN(CA)NNH	156(^{15}N)	156(^{15}N)	2048(^{1}H)	26	26	15	24	0.25
TROSY HN(COCA)NNH	96(^{15}N)	150(^{15}N)	2048(^{1}H)	26	26	15	16	1
Structure and dynamics								
^{15}N R_1^d	512(^{15}N)	2048(^{1}H)		24	16.1		8	3.0
^{15}N R_2^d	512(^{15}N)	2048(^{1}H)		24	16.1		8	3.0
Steady-state heteronuclear ^{1}H-^{15}N NOEs[d]	1024(^{15}N)	2048(^{1}H)		24	16.1		32	6.0

[a] Number of acquired scans.

[b] Relaxation delay in seconds.

[c] For experiments acquired in the IPAP mode, the dimension in which the two experiments are stored is indicated with an asterisk.

[d] In ^{15}N R_1, ^{15}N R_2 and heteronuclear ^{1}H-^{15}N-NOEs experiments the water signal was suppressed with "water flip-back" scheme. For the determination of R_1, 10 experiments were acquired changing the variable delay from 20 ms to 850 ms. For the determination of R_2, 9 experiments were acquired changing the variable delay from 18 ms to 250 ms. All 3D and 2D spectra were processed using the standard Bruker software (TopSpin 2.0) and analyzed through the ccpNMR (Collaborative Computing Project for NMR) [13].

Table S2: Experiments used for HAdV 2 E1A-12S sequence specific assignment.

Experiments used for HAdV 2 E1A-12S sequence specific assignment	**Dimension of acquired data**			**Spectral width (ppm)**			n[a]	d[b]
	t_1	t_2	t_3	F_1	F_2	F_3		
^{13}C detected								
CON-IPAP[c]	1024*(^{15}N)	1024(^{13}C)		40	30		24	2
^{1}H detected								
^{1}H-^{15}N BEST-TROSY	512(^{15}N)	2048(^{1}H)		23	12		4	0.2
BEST-TROSY HNCO	164(^{13}C)	196(^{15}N)	2048(^{1}H)	9.5	23	15	4	0.2
BEST-TROSY HN(CA)CO	164(^{13}C)	196(^{15}N)	2048(^{1}H)	9.5	23	15	8	0.2
BEST-TROSY HN(CO)CACB	188(^{13}C)	196(^{15}N)	2048(^{1}H)	58.1	23	15	8	0.2
BEST-TROSY HNCACB	188(^{13}C)	196(^{15}N)	2048(^{1}H)	58.1	23	15	8	0.2
BEST-TROSY HN(CA)NNH	138(^{15}N)	138(^{15}N)	2048(^{1}H)	23	23	15	24	0.28
TROSY HN(COCA)NNH	104(^{15}N)	168(^{15}N)	2048(^{1}H)	23	23	15	8	1
Structure and dynamics								
^{15}N R_1[d]	350(^{15}N)	2048(^{1}H)		24	15		16	3.0
^{15}N R_2[d]	350(^{15}N)	2048(^{1}H)		24	15		16	3.0
Steady-state heteronuclear ^{1}H-^{15}N NOEs[d]	640(^{15}N)	2048(^{1}H)		24	15		48	6.0

[a] Number of acquired scans.

[b] Relaxation delay in seconds.

[c] For experiments acquired in the IPAP mode, the dimension in which the two experiments are stored is indicated with an asterisk.

[d] In ^{15}N R_1, ^{15}N R_2 and heteronuclear ^{1}H-^{15}N-NOEs experiments the water signal was suppressed with "water flip-back" scheme. For the determination of R_1, 10 experiments were acquired changing the variable delay from 20 ms to 1 s. For the determination of R_2, 10 experiments were acquired changing the variable delay from 18 ms to 250 ms.
All 3D and 2D spectra were processed using the standard Bruker software (TopSpin 2.0) and analyzed through the ccpNMR (Collaborative Computing Project for NMR) [13].

Table S3: 4D experiments used for HAdV 2 E1A-12S

Experiments used for HAdV 2 E1A-12S	**Spectral width and maximal evolution times**				n[a]	d[b]	cp[c]	hp[d]	t^{e}_{exp}	P[f]
^{13}C detected										
HCBCACON	8800 Hz	2600 Hz	12500 Hz	5000 Hz	8	0.9	512	850	1 day, 11 h	0.07
	(^{13}C)	(^{15}N)	(^{13}C^{ali})	(^{1}H)						
		50 ms	7.6 ms	10 ms						
HCBCANCO	8800 Hz	2600 Hz	12500 Hz	5000 Hz	16	0.9	512	850	3 day, 2 h	0.11
	(^{13}C)	(^{15}N)	(^{13}C^{ali})	(^{1}H)						
		32 ms	7.6 ms	10 ms						
HCACONCACON	8800 Hz	2600 Hz	2200 Hz	2600 Hz	16	0.9	512	930	3 day, 10 h	0.34
	(^{13}C)	(^{15}N)	(^{13}C')	(^{15}N)						
		30 ms	25 ms	25 ms						
HNCONCACON	8800 Hz	2600 Hz	2200 Hz	2600 Hz	16	0.5	512	910	2 day, 6 h	0.33
	(^{13}C)	(^{15}N)	(^{13}C')	(^{15}N)						
		30 ms	25 ms	25 ms						

[a] number of acquired scans.
[b] inter-scan delay (s).
[c] number of complex points.
[d] number of hypercomplex points.
[e] duration of the experiment.
[f] relative data points density (%).

Figure S1. The prediction of disordered regions was done with the IUPred [14] and combined with the prediction of binding regions by the ANCHOR [15] of HAdV 2 E1A-12S (**a**) and HAdV 2/5 E1A-13S (**b**). **c** The predictions of disorder done with the metapredictor PONDR-FIT [16] for HAdV 2 E1A-12S (red) and HAdV 2/5 E1A-13S (blue) protein sequences are compared. The alignment between both isoforms was done taking in consideration the 46 residues of the CR3 region between residues 144 and 191 of the E1A-13S missing in E1A-12S.

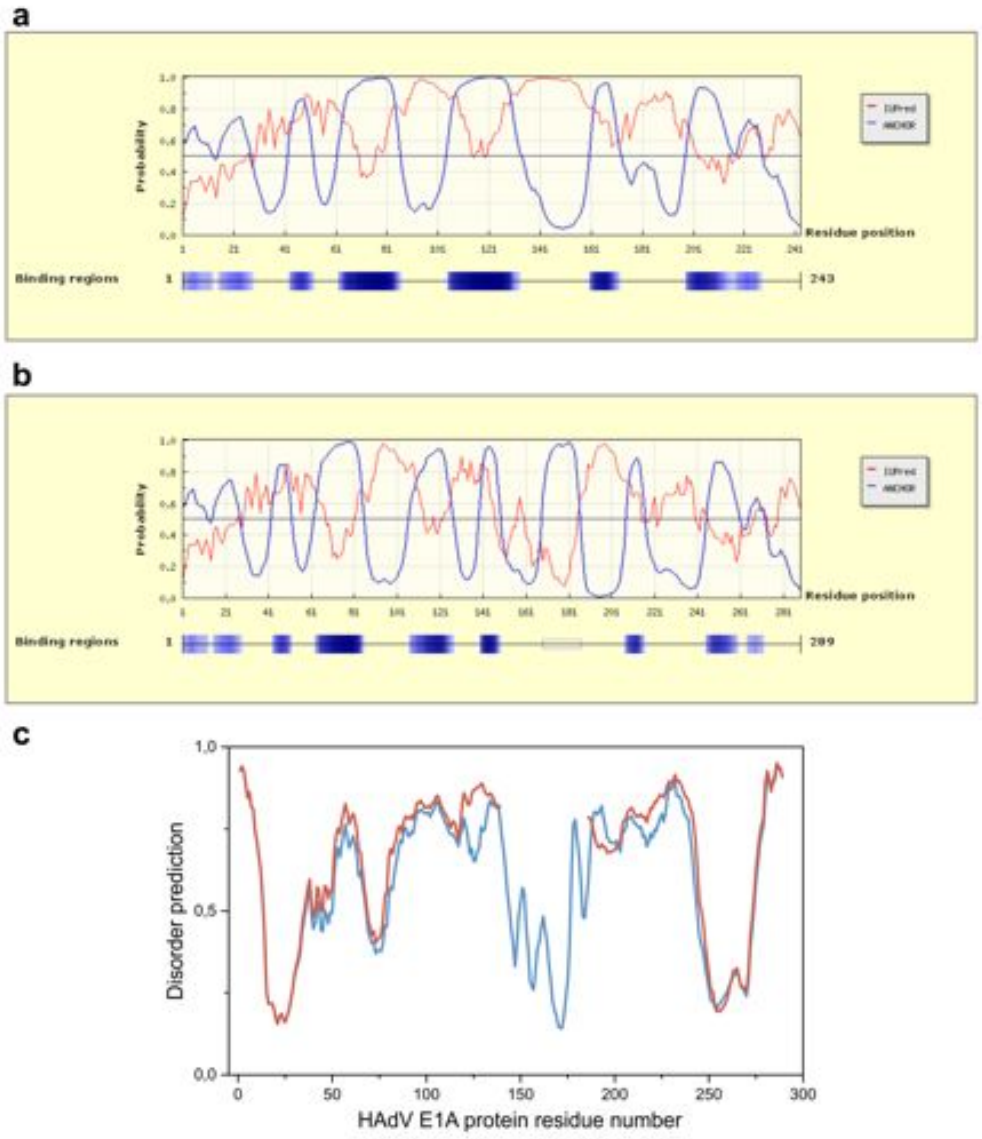
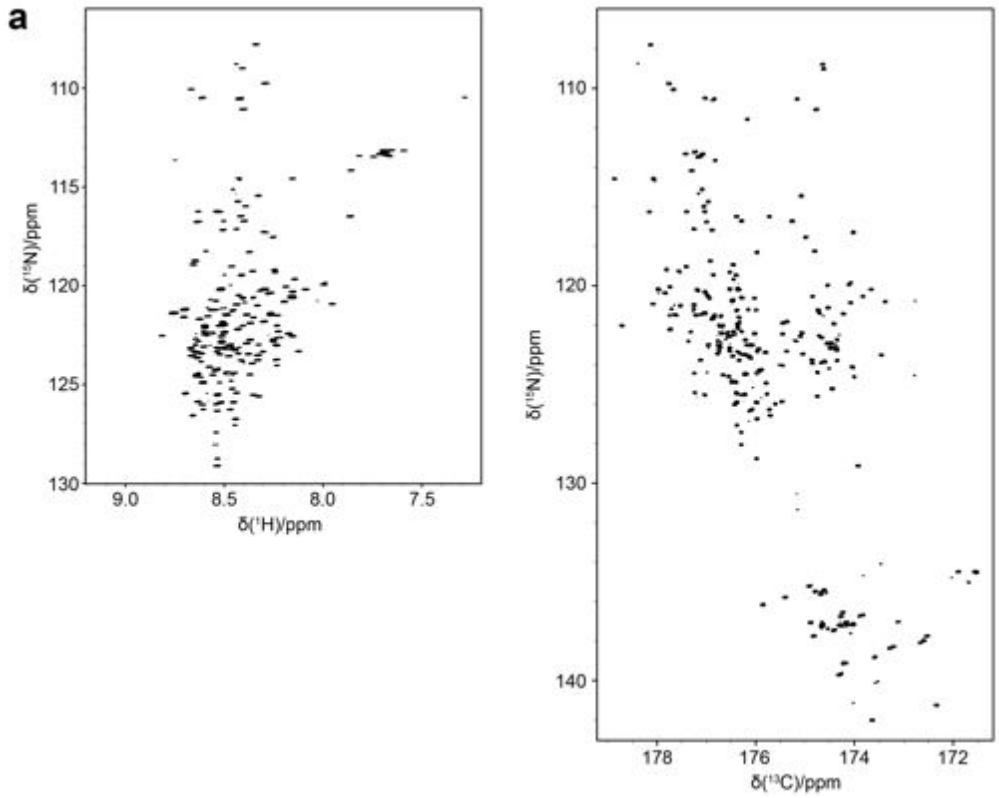
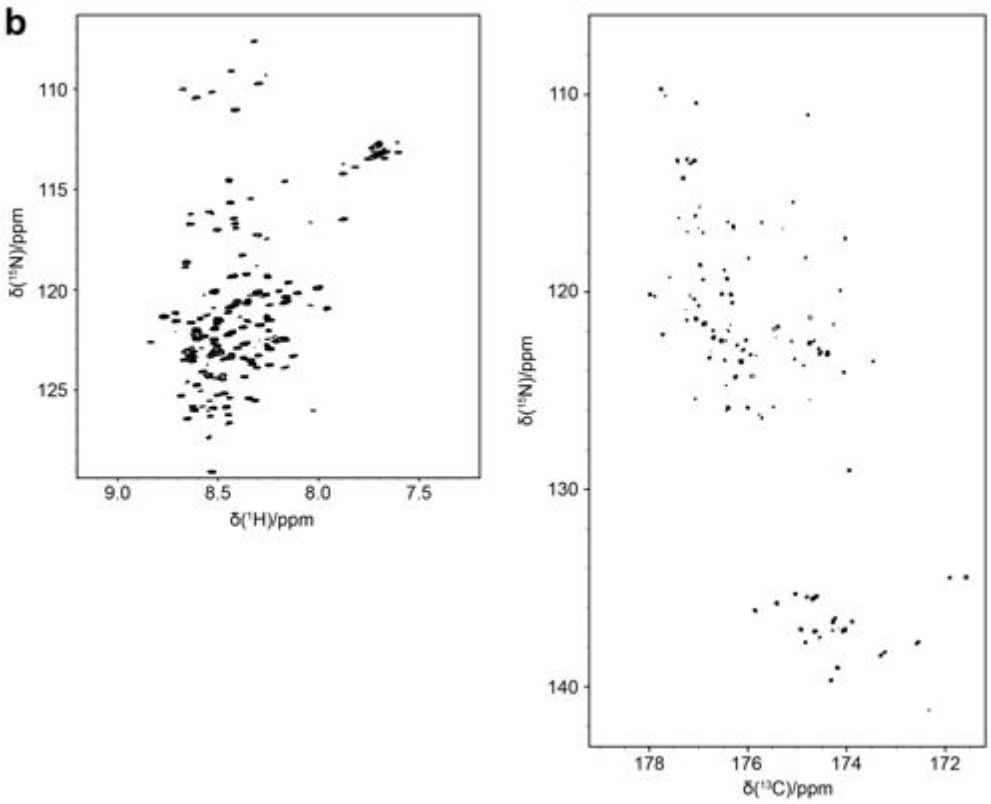


Figure S2. ^{1}H-^{15}N HSQC and ^{13}C-^{15}N CON-IPAP spectra recorded for HAdV 2 E1A-12S (**a**) and HAdV 2/5 E1A-13S (**b**). Experiments were recorded at 16.4 T spectrometer and 278 K. The conditions of the HAdV 2/5 E1A-13S sample were 0.5 mM protein concentration in 10 mM HEPES buffer pH 7.5, 50 mM KCl and 10 mM DTT. The protein concentration of the HAdV 2/5 E1A-13S sample was 0.2 mM in 10 mM HEPES buffer pH 7.5, 150 mM KCl, 10 μM $ZnCl_2$, and 10 mM DTT.




b
110
115
120
125
δ(15N)/ppm
9.0
8.5
8.0
7.5
δ(1H)/ppm
110
120
130
140
δ(15N)/ppm
178
176
174
172
δ(13C)/ppm

References

[1] Bermel W, Bertini I, Duma L, Felli IC, Emsley L, Pierattelli R, Vasos PR. Complete Assignment of Heteronuclear Protein Resonances by Protonless NMR Spectroscopy. *Angew Chem Int Ed.* 2005;44(20):3089–3092.

[2] Bermel W, Bertini I, Felli IC, Gonnelli L, Koźmiński W, Piai A, Pierattelli R, Stanek J. Speeding up sequence specific assignment of IDPs. *J Biomol NMR.* 2012;53(4):293–301.

[3] Bermel W, Felli IC, Gonnelli L, Koźmiński W, Piai A, Pierattelli R, Zawadzka-Kazimierczuk A. High-dimensionality C-13 direct-detected NMR experiments for the automatic assignment of intrinsically disordered proteins. *J Biomol NMR.* 2013;57(4):353–361.

[4] Favier A, Brutscher B. Recovering lost magnetization: polarization enhancement in biomolecular NMR. *J Biomol NMR.* 2011;49(1):9–15.

[5] Lescop E, Kern T, Brutscher B. Guidelines for the use of band-selective radiofrequency pulses in hetero-nuclear NMR: example of longitudinal-relaxation-enhanced BEST-type 1H-15N correlation experiments. *J Magn Reson.* 2010;203(1):190–198.

[6] Lescop E, Schanda P, Brutscher B. A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *Journal of Magnetic Resonance.* 2007;187(1):163–169.

[7] Salzmann MM, Pervushin K, Wider GG, Senn HH, Wüthrich K. TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. *Proc Natl Acad Sci USA.* 1998;95(23):13585–13590.

[8] Solyom Z, Schwarten M, Geist L, Konrat R, Willbold D, Brutscher B. BEST-TROSY experiments for time-efficient sequential resonance assignment of large disordered proteins. *J Biomol NMR.* 2013;55(4):311–321.

[9] Schanda P, Van Melckebeke H, Brutscher B. Speeding up three-dimensional protein NMR experiments to a few minutes. *J Am Chem Soc.* 2006;128(28):9042–9043.

[10] Weisemann R, Rüterjans H, Bermel W. 3D Triple-resonance NMR techniques for the sequential assignment of NH and 15N resonances in 15N- and 13C-labelled proteins. *J Biomol NMR.* 1993;3(1):113–120.

[11] Bracken C, Palmer AG III, Cavanagh J. (H)N(COCA)NH and HN(COCA)NH experiments for 1H-15N backbone assignments in 13C/15N-labeled proteins. *J Biomol NMR.* 1997;9(1):94–100.

[12] Panchal SC, Bhavesh NS, Hosur RV. Improved 3D triple resonance experiments, HNN and HN(C)N, for HN and 15N sequential correlations in (13C, 15N) labeled proteins: application to unfolded proteins. *J Biomol NMR.* 2001;20(2):135–147.

[13] Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J, Laue ED. The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins*. 2005;59(4):687–696.

[14] Dosztányi Z, Csizmok V, Tompa P, Simon I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics*. 2005;21(16):3433–3434.

[15] Dosztányi Z, Mészáros B, Simon I. ANCHOR: web server for predicting protein binding regions in disordered proteins. *J Gerontol*. 2009;25(20):2745–2746.

[16] Xue B, Dunbrack RL, Williams RW, Dunker AK, Uversky VN. PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochim Biophys Acta*. 2010;1804(4):996–1010.

Chapter 6

Conclusions and Perspectives

Intrinsically disordered proteins have expanded our idea of the mechanisms behind protein function. Thanks to their peculiar properties they are involved in many key regulatory and signaling processes in cells. Hence their detailed characterization is essential for the understanding of key cellular mechanisms, whose malfunction causes serious diseases. Among the biophysical techniques NMR spectroscopy occupies a unique position since it is able to provide high-resolution information also on highly dynamic macromolecules such as IDPs. Nonetheless, the development of the NMR methods has for a long time focused on the investigation of structured proteins, as a result of the influence of the structural-function paradigm. NMR experiments optimized for the study of folded proteins are not necessarily also optimal IDPs. This has stimulated the development of NMR experiments tailored for the characterization of IDPs, which is the central topic of this thesis.

The identification of the spectroscopic challenges originating from the highly dynamic nature of IDPs is the initial step towards the optimization and design of new NMR methods aiming at their investigation. The lack of a stable 3D structure has both advantages and disadvantages from the spectroscopic point of view. The high flexibility, in in case of IDPs amenable to NMR, results in short rotational correlation times and thus slowly relaxing spin coherences, which allows the design of complicated pulse sequence schemes. On the other hand, the low chemical shift dispersion and fast exchange of labile protons with solvent protons responsible for line-broadening represent problems that should be addressed. The common aspects of the experiments proposed are the use of LRE techniques and the exploitation of the superior chemical shift dispersion of backbone carbonyls and nitrogens.

Fast hydrogen exchange rates, provided they lie in a specific regime, can be exploited to significantly improve experimental sensitivity by combining LRE techniques with ^{13}C de-

tection. This approach, employed for the design of the $H^{N\text{-}BEST}CON$ pulse sequence, is particularly useful for monitoring of post-translational modifications at close to physiological conditions and for studying the behavior of IDPs inside living cells. The selective manipulation of different pools of proton spins, at the basis of LRE, adopted for the design of IR-$H^{N\text{-}flip}CON$ and IR-$H^{\alpha\text{-}BEST}CON$ experiments, provides information on overall longitudinal proton relaxation properties. This provides interesting information on the local proton density and solvent exchange processes, which report on the local structural and dynamic properties of a protein. The investigation of longitudinal relaxation in IDPs is also important for the design of new experiments that exploit ^1H polarization as a starting source. Indeed these findings were used for the design of additional 5D multidimensional experiments exploiting C'_i-N_{i+1} pairs for sequence-specific assignment, either based on direct detection of $^1H^N$ or $^1H^\alpha$. These 5D NMR experiments, together with the previously developed ones based on ^{13}C detection, now form a complete suite to assign IDPs ("CON-CON" assignment strategy).

NMR experiments optimized for studying IDPs resulted particularly useful for the characterization of two viral IDPs: E7 from HPV 16 and E1A from HAdV 2/5. Both proteins play a crucial, however opposite, role in cancer development in humans, yet neither of them had been investigated by high-resolution techniques in their full-length form. Our NMR studies revealed very heterogeneous structural and dynamic properties of E7, displaying variable extent of disorder going beyond the schematic view of a fully disordered N-terminal half and a well-structured CR3 region. Similar heterogeneous structural and dynamic properties were also identified for E1A. In this case investigation confirmed the presence of disordered, flexible parts comprising the CR1, CR2 and CR4 regions, while the CR3 region escaped detection through NMR. Its more ordered, structured nature was verified by other biophysical techniques.

In conclusion, the development of the new NMR methods in parallel with the improvements of the instrumental sensitivity extends the number of IDPs that can be studied at atomic resolution by NMR. The proposed experiments represent a promising tool for the investigation of IDPs, in particular approaching physiological conditions. The simple 2D experiments proposed based on carbonyl direct detection provide a useful tool for *in-cell* applications, as they can provide additional experimental information on the structural and dynamic behavior of the IDPs inside the cell that are difficult to access by other techniques. Higher dimensionality experiments are the key to extend the size and complexity of IDPs that can be characterized at atomic resolution, Furthermore, the thorough NMR characterization of the two viral IDPs has laid the basis for future high-resolution studies

of their complex interaction network and the effect of the post-translational modification on their function. These investigations provide important insights into the molecular mechanisms that enable IDPs to act as hub proteins and thus interfere with many cellular processes.

Finally, speculating on more long term perspectives, development of improved NMR methods to study IDPs is expected to provide a large amount of experimental data on IDPs contributing to our understanding of the molecular basis responsible for their function, filling a gap of about 50 years respect to our knowledge on the structural and dynamic behavior of folded proteins, well documented by the many data available in the PDB. This is expected to reveal a much larger number of ways in which proteins communicate in the cell. Other expected outcomes of NMR experimental data on IDPs include the improvement of prediction tools, which are now largely based on missing electron density maps in X-ray crystallography.

Bibliography

[1] Garner E, Cannon P, Romero P, Obradovic Z, Dunker AK. Predicting Disordered Regions from Amino Acid Sequence: Common Themes Despite Differing Structural Characterization. *Genome Inform Ser Workshop Genome Inform*. 1998;9:201–213.

[2] Romero P, Obradovic Z, Kissinger CR, Villafranca JE, Garner E, Guilliot S, Dunker AK. Thousands of proteins likely to have long disordered regions. *Pac Symp Biocomput*. 1998;pp. 437–448.

[3] Abercrombie BD, Kneale GG, Crane-Robinson C, Bradbury EM, Goodwin GH, Walker JM, Johns EW. Studies on the conformational properties of the high-mobility-group chromosomal protein HMG 17 and its interaction with DNA. *Eur J Biochem*. 1978;84(1):173–177.

[4] Cary PD, King DS, Crane-Robinson C, Bradbury EM, Rabbani A, Goodwin GH, Johns EW. Structural studies on two high-mobility-group proteins from calf thymus, HMG-14 and HMG-20 (ubiquitin), and their interaction with DNA. *Eur J Biochem*. 1980;112(3):577–580.

[5] Cary PD, Crane-Robinson C, Bradbury EM, Dixon GH. Structural studies of the non-histone chromosomal proteins HMG-T and H6 from trout testis. *Eur J Biochem*. 1981;119(3):545–551.

[6] Hernández MA, Avila J, Andreu JM. Physicochemical characterization of the heat-stable microtubule-associated protein MAP2. *Eur J Biochem*. 1986;154(1):41–48.

[7] Lynch WP, Riseman VM, Bretscher A. Smooth muscle caldesmon is an extended flexible monomeric protein in solution that can readily undergo reversible intra- and intermolecular sulfhydryl cross-linking. A mechanism for caldesmon's F-actin bundling activity. *J Biol Chem*. 1987;262(15):7429–7437.

[8] Thomas J, Van Patten SM, Howard P, Day KH, Mitchell RD, Sosnick T, Trewhella J, Walsh DA, Maurer RA. Expression in Escherichia coli and characterization of the heat-stable inhibitor of the cAMP-dependent protein kinase. *J Biol Chem.* 1991; 266(17):10906–10911.

[9] Schweers O, Schönbrunn-Hanebeck E, Marx A, Mandelkow E. Structural studies of tau protein and Alzheimer paired helical filaments show no evidence for beta-structure. *J Biol Chem.* 1994;269(39):24290–24297.

[10] Cho HS, Liu CW, Damberger FF, Pelton JG, Nelson HC, Wemmer DE. Yeast heat shock transcription factor N-terminal activation domains are unstructured as probed by heteronuclear NMR spectroscopy. *Protein Sci.* 1996;5(2):262–269.

[11] Bhattacharyya J, Das KP. Molecular chaperone-like properties of an unfolded protein, alpha(s)-casein. *J Biol Chem.* 1999;274(22):15505–15509.

[12] Uversky VN, Gillespie JR, Fink AL. Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins.* 2000;41(3):415–427.

[13] Vacic V, Oldfield CJ, Mohan A, Radivojac P, Cortese MS, Uversky VN, Dunker AK. Characterization of molecular recognition features, MoRFs, and their binding partners. *J Proteome Res.* 2007;6(6):2351–2366.

[14] Sigler PB. Transcriptional activation. Acid blobs and negative noodles. *Nature.* 1988;333(6170):210–212.

[15] Dobson CM. Flexible friends. *Curr Biol.* 1993;3(8):530–532.

[16] Dunker AK, Garner E, Guilliot S, Romero P, Albrecht K, Hart J, Obradovic Z, Kissinger C, Villafranca JE. Protein disorder and the evolution of molecular recognition: theory, predictions and observations. *Pac Symp Biocomput.* 1998;pp. 473–484.

[17] Dunker AK, Obradovic Z, Romero P, Garner EC. Intrinsic protein disorder in complete genomes. *Genome Informatics.* 2000;11:161–171.

[18] Wright PE, Dyson HJ. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *Journal of Molecular Biology.* 1999;293(2):321–331.

[19] Tompa P. Intrinsically unstructured proteins. *Trends in Biochemical Sciences.* 2002; 27(10):527–533.

[20] Uversky VN. Intrinsically disordered proteins from A to Z. *The International Journal of Biochemistry & Cell Biology*. 2011;43(8):1090–1103.

[21] Sickmeier M, Hamilton JA, LeGall T, Vacic V, Cortese MS, Tantos A, Szabo B, Tompa P, Chen J, Uversky VN, Obradovic Z, Dunker AK. DisProt: the Database of Disordered Proteins. *Nucleic Acids Research*. 2007;35:D786–93.

[22] Varadi M, Kosol S, Lebrun P, Valentini E, Blackledge M, Dunker AK, Felli IC, Forman-Kay JD, Kriwacki RW, Pierattelli R, Sussman J, Svergun DI, Uversky VN, Vendruscolo M, Wishart D, Wright PE, Tompa P. pE-DB: a database of structural ensembles of intrinsically disordered and of unfolded proteins. *Nucleic Acids Research*. 2014;42(Database issue):D326–35.

[23] Ozenne V, Bauer F, Salmon L, Huang JR, Jensen MR, Segard S, Bernadó P, Charavay C, Blackledge M. Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. *Bioinformatics*. 2012;28(11):1463–1470.

[24] Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol*. 2005;6(3):197–208.

[25] Peng K, Radivojac P, Vucetic S, Dunker AK, Obradovic Z. Length-dependent prediction of protein intrinsic disorder. *BMC Bioinformatics*. 2006;7:208.

[26] Uversky VN. A decade and a half of protein intrinsic disorder: biology still waits for physics. *Protein Science*. 2013;22(6):693–724.

[27] Uversky VN, Dunker AK. Understanding protein non-folding. *Biochim Biophys Acta*. 2010;1804(6):1231–1264.

[28] Tompa P. The interplay between structure and function in intrinsically unstructured proteins. *FEBS Letters*. 2005;579(15):3346–3354.

[29] Romero PR, Zaidi S, Fang YY, Uversky VN, Radivojac P, Oldfield CJ, Cortese MS, Sickmeier M, LeGall T, Obradovic Z, Dunker AK. Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms. *Proc Natl Acad Sci USA*. 2006;103(22):8390–8395.

[30] Iakoucheva LM, Brown CJ, Obradovic Z, Dunker AK, Lawson JD. Intrinsic disorder in cell-signaling and cancer-associated proteins. *Journal of Molecular Biology*. 2002; 323(3):573–584.

[31] Xie H, Vucetic S, Iakoucheva LM, Dunker AK, Oldfield CJ, Uversky VN, Obradovic Z. Functional Anthology of Intrinsic Disorder. 1. Biological Processes and Functions of Proteins with Long Disordered Regions. *J Proteome Res.* 2007;6:1882–1898.

[32] Vucetic S, Xie H, Iakoucheva LM, Oldfield CJ, Dunker AK, Obradovic Z, Uversky VN. Functional Anthology of Intrinsic Disorder. 2. Cellular Components, Domains, Technical Terms, Developmental Processes, and Coding Sequence Diversities Correlated with Long Disordered Regions. *J Proteome Res.* 2007;6(5):1899–1916.

[33] Jeffery CJ. Moonlighting proteins. *Trends in Biochemical Sciences.* 1999;24(1):8–11.

[34] Tompa P, Szász C, Buday L. Structural disorder throws new light on moonlighting. *Trends in Biochemical Sciences.* 2005;30(9):484–489.

[35] Barabasi AL, Albert R. Emergence of scaling in random networks. *Science.* 1999; 286(5439):509–512.

[36] Goh KI, Oh E, Jeong H, Kahng B, Kim D. Classification of scale-free networks. *Proc Natl Acad Sci USA.* 2002;99(20):12583–12588.

[37] Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN. Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS Journal.* 2005; 272(20):5129–5148.

[38] Patil A, Nakamura H. Disordered domains and high surface charge confer hubs with the ability to interact with multiple proteins in interaction networks. *FEBS Letters.* 2006;580(8):2041–2045.

[39] Dosztányi Z, Chen J, Dunker AK, Simon I, Tompa P. Disorder and sequence repeats in hub proteins and their implications for network evolution. *J Proteome Res.* 2006; 5(11):2985–2995.

[40] Radivojac P, Iakoucheva LM, Oldfield CJ, Obradovic Z, Uversky VN, Dunker AK. Intrinsic disorder and functional proteomics. *Biophys J.* 2007;92(5):1439–1456.

[41] Kalderon D, Roberts BL, Richardson WD, Smith AE. A short amino acid sequence able to specify nuclear location. *Cell.* 1984;39(3):499–509.

[42] Gould CM, Diella F, Via A, Puntervoll P, Gemünd C, Chabanis-Davidson S, Michael S, Sayadi A, Bryne JC, Chica C, Seiler M, Davey NE, Haslam N, Weatheritt RJ, Budd A, Hughes T, Pas J, Rychlewski L, Travé G, Aasland R, Helmer-Citterich M,

Linding R, Gibson TJ. ELM: the status of the 2010 eukaryotic linear motif resource. *Nucleic Acids Research*. 2010;38:D167–D180.

[43] Tompa P, Fuxreiter M. Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends in Biochemical Sciences*. 2008;33(1):2–8.

[44] Fuxreiter M, Tompa P. FUZZY COMPLEXES: A More Stochastic View of Protein Function. *Adv Exp Med Biol*. 2012;725:1–14.

[45] Uversky VN. Multitude of binding modes attainable by intrinsically disordered proteins: a portrait gallery of disorder-based complexes. *Chem Soc Rev*. 2011; 40(3):1623–1634.

[46] Kelly JW. The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways. *Current Opinion in Structural Biology*. 1998;8(1):101–106.

[47] Bellotti V, Mangione P, Stoppini M. Biological activity and pathological implications of misfolded proteins. *Cellular and Molecular Life Sciences (CMLS)*. 1999;55(6-7):977–991.

[48] Uversky VN, Fink AL. Conformational constraints for amyloid fibrillation: the importance of being unfolded. *Biochim Biophys Acta*. 2004;1698(2):131–153.

[49] Masters CL, Multhaup G, Simms G, Pottgiesser J, Martins RN, Beyreuther K. Neuronal Origin of a Cerebral Amyloid - Neurofibrillary Tangles of Alzheimers-Disease Contain the Same Protein as the Amyloid of Plaque Cores and Blood-Vessels. *EMBO J*. 1985;4(11):2757–2763.

[50] Dev KK, Hofele K, Barbieri S, Buchman VL, van der Putten H. Part II: alpha-synuclein and its molecular pathophysiological role in neurodegenerative disease. *Neuropharm*. 2003;45(1):14–44.

[51] Hollstein M, Sidransky D, Vogelstein B, Harris CC. P53 Mutations in Human Cancers. *Science*. 1991;253(5015):49–53.

[52] Balint E, Vousden KH. Activation and activities of the p53 tumour suppressor protein. *Br J Cancer*. 2001;85(12):1813–1823.

[53] Uversky VN, Oldfield CJ, Dunker AK. Intrinsically Disordered Proteins in Human Diseases: Introducing the D 2Concept. *Annu Rev Biophys*. 2008;37(1):215–246.

[54] Davey NE, Travé G, Gibson TJ. How viruses hijack cell regulation. *Trends in Biochemical Sciences*. 2011;36(3):159–169.

[55] Parkin DM. The global health burden of infection -associated cancers in the year 2002. *Int J Cancer*. 2006;118(12):3030–3044.

[56] Frisch SM, Mymryk JS. Adenovirus-5 e1a: paradox and paradigm. *Nat Rev Mol Cell Biol*. 2002;3(6):441–452.

[57] Bonvin AMJJ, Boelens R, Kaptein R. NMR analysis of protein interactions. *Curr Opin Chem Biol*. 2005;9(5):501–508.

[58] Kaptein R, Boelens R, Scheek RM, van Gunsteren WF. Protein structures from NMR. *Biochemistry*. 1988;27(15):5389–5395.

[59] Schanda P, Forge V, Brutscher B. HET-SOFAST NMR for fast detection of structural compactness and heterogeneity along polypeptide chains. *Magn Reson Chem*. 2006;44(S1):S177–S184.

[60] Rennella E, Solyom Z, Brutscher B. Measuring hydrogen exchange in proteins by selective water saturation in (1)H- (15)N SOFAST/BEST-type experiments: advantages and limitations. *J Biomol NMR*. 2014;60(2-3):99–107.

[61] Theillet FX, Kalmar L, Tompa P, Han KH, Selenko P, Dunker AK, Daughdrill GW, Uversky VN. The alphabet of intrinsic disorder. *Intrinsically Disord Proteins*. 2014; 1(1):e24360.

[62] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. *Nucleic Acids Research*. 2000;28(1):235–242.

[63] Wirmer J, Berk H, Ugolini R, Redfield C, Schwalbe H. Characterization of the unfolded state of bovine alpha-lactalbumin and comparison with unfolded states of homologous proteins. *Protein Sci*. 2006;15(6):1397–1407.

[64] Gerum C, Silvers R, Wirmer-Bartoschek J, Schwalbe H. Unfolded-state structure and dynamics influence the fibril formation of human prion protein. *Angew Chem Int Ed*. 2008;48(50):9452–9456.

[65] Felli IC, Pierattelli R. Novel methods based on 13C detection to study intrinsically disordered proteins. *Journal of Magnetic Resonance*. 2014;241:115–125.

[66] Bayro MJ, Maly T, Birkett NR, Macphee CE, Dobson CM, Griffin RG. High-resolution MAS NMR analysis of PI3-SH3 amyloid fibrils: backbone conformation and implications for protofilament assembly and structure . *Biochemistry*. 2010; 49(35):7474–7484.

[67] Bayro MJ, Debelouchina GT, Eddy MT, Birkett NR, MacPhee CE, Rosay M, Maas WE, Dobson CM, Griffin RG. Intermolecular structure determination of amyloid fibrils with magic-angle spinning and dynamic nuclear polarization NMR. *J Am Chem Soc*. 2011;133(35):13967–13974.

[68] Bertini I, Gallo G, Korsak M, Luchinat C, Mao J, Ravera E. Formation kinetics and structural features of Beta-amyloid aggregates by sedimented solute NMR. *ChemBioChem*. 2013;14(14):1891–1897.

[69] Schanda P, Van Melckebeke H, Brutscher B. Speeding up three-dimensional protein NMR experiments to a few minutes. *J Am Chem Soc*. 2006;128(28):9042–9043.

[70] Lescop E, Schanda P, Brutscher B. A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *Journal of Magnetic Resonance*. 2007;187(1):163–169.

[71] Bermel W, Bertini I, Felli IC, Pierattelli R. Speeding Up 13C Direct Detection Biomolecular NMR Spectroscopy. *J Am Chem Soc*. 2009;131(42):15339–15345.

[72] Jaravine VA, Zhuravleva AV, Permi P, Ibraghimov I, Orekhov VY. Hyperdimensional NMR Spectroscopy with Nonlinear Sampling. *J Am Chem Soc*. 2008; 130(12):3927–3936.

[73] Kazimierczuk K, Zawadzka A, Koźmiński W, Zhukov I. Random sampling of evolution time space and Fourier transform processing. *J Biomol NMR*. 2006;36(3):157–168.

[74] Liu A, Riek R, Wider G, von Schroetter C, Zahn R, Wüthrich K. NMR experiments for resonance assignments of 13C, 15N doubly-labeled flexible polypeptides: application to the human prion protein hPrP(23-230). *J Biomol NMR*. 2000;16(2):127–138.

[75] Löhr F, Pfeiffer S, Lin YJ, Hartleib J, Klimmek O, Rüterjans H. HNCAN pulse sequences for sequential backbone resonance assignment across proline residues in perdeuterated proteins. *J Biomol NMR*. 2000;18(4):337–346.

[76] Schubert M, Ball LJ, Oschkinat H, Schmieder P. Bridging the gap: A set of selective 1H-15N-correlations to link sequential neighbors of prolines. *J Biomol NMR*. 2000; 17(4):331–335.

[77] Solyom Z, Schwarten M, Geist L, Konrat R, Willbold D, Brutscher B. BEST-TROSY experiments for time-efficient sequential resonance assignment of large disordered proteins. *J Biomol NMR*. 2013;55(4):311–321.

[78] Kanelis V, Donaldson L, Muhandiram DR, Rotin D, Forman-Kay JD, Kay LE. Sequential assignment of proline-rich regions in proteins: application to modular binding domain complexes. *J Biomol NMR*. 2000;16(3):253–259.

[79] Mäntylahti S, Aitio O, Hellman M, Permi P. HA-detected experiments for the backbone assignment of intrinsically disordered proteins. *J Biomol NMR*. 2010; 47(3):171–181.

[80] Bertini I, Duma L, Felli IC, Fey M, Luchinat C, Pierattelli R, Vasos PR. A Heteronuclear Direct-Detection NMR Spectroscopy Experiment for Protein-Backbone Assignment. *Angew Chem Int Ed*. 2004;43(17):2257–2259.

[81] Csizmok V, Felli IC, Tompa P, Banci L, Bertini I. Structural and Dynamic Characterization of Intrinsically Disordered Human Securin by NMR Spectroscopy. *J Am Chem Soc*. 2008;130(50):16873–16879.

[82] Salmon L, Jensen MR, Bernadó P, Blackledge M. Measurement and analysis of NMR residual dipolar couplings for the study of intrinsically disordered proteins. *Methods Mol Biol*. 2011;895:115–125.

[83] Wu KP, Baum J. Detection of transient interchain interactions in the intrinsically disordered protein alpha-synuclein by NMR paramagnetic relaxation enhancement. *J Am Chem Soc*. 2010;132(16):5546–5547.

[84] Schwalbe H, Carlomagno T, Hennig M, Junker J, Reif B, Richter C, Griesinger C. Cross-correlated relaxation for measurement of angles between tensorial interactions. *Methods Enzymol*. 2000;338:35–81.

[85] Pervushin K, Riek R, Wider G, Wüthrich K. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci USA*. 1997;94(23):12366–12371.

[86] Schanda P, Kupče E, Brutscher B. SOFAST-HMQC Experiments for Recording Two-dimensional Deteronuclear Correlation Spectra of Proteins within a Few Seconds. *J Biomol NMR*. 2005;33(4):199–211.

[87] Mäntylahti S, Hellman M, Permi P. Extension of the HA-detection based approach: (HCA)CON(CA)H and (HCA)NCO(CA)H experiments for the main-chain assignment of intrinsically disordered proteins. *J Biomol NMR*. 2011;49(2):99–109.

[88] Emsley L, Bodenhausen G. Phase shifts induced by transient Bloch-Siegert effects in NMR. *Chem Phys Lett*. 1990;168(3):297–303.

[89] Waugh JS. Theory of broadband spin decoupling. *J Magn Reson*. 1982;50(1):30–49.

[90] Ying J, Li F, Lee JH. 13Cα decoupling during direct observation of carbonyl resonances in solution NMR of isotopically enriched proteins. *J Biomol NMR*. 2014; 60(1):15–21.

[91] Shimba N, Stern AS, Craik CS, Hoch JC, Dötsch V. Elimination of 13Calpha splitting in protein NMR spectra by deconvolution with maximum entropy reconstruction. *J Am Chem Soc*. 2003;125(9):2382–2383.

[92] Bermel W, Felli IC, Kümmerle R, Pierattelli R. 13C Direct-detection biomolecular NMR. *Concepts Magn Reson*. 2008;32A(3):183–200.

[93] Meissner A, Duus JO, Sorensen OW. Spin-state-selective excitation. Application for E.COSY-type measurement of J(HH) coupling constants. *Journal of Magnetic Resonance*. 1997;128(1):92–97.

[94] Bermel W, Bertini I, Duma L, Felli IC, Emsley L, Pierattelli R, Vasos PR. Complete Assignment of Heteronuclear Protein Resonances by Protonless NMR Spectroscopy. *Angew Chem Int Ed*. 2005;44(20):3089–3092.

[95] Kern T, Schanda P, Brutscher B. Sensitivity-enhanced IPAP-SOFAST-HMQC for fast-pulsing 2D NMR with reduced radiofrequency load. *J Magn Reson*. 2008; 190(2):333–338.

[96] Bermel W, Bertini I, Csizmok V, Felli IC, Pierattelli R, Tompa P. H-start for exclusively heteronuclear NMR spectroscopy: The case of intrinsically disordered proteins. *Journal of Magnetic Resonance*. 2009;198(2):275–281.

[97] Gil S, Hošek T, Solyom Z, Kümmerle R, Brutscher B, Pierattelli R, Felli IC. NMR Spectroscopic Studies of Intrinsically Disordered Proteins at Near-Physiological Conditions. *Angew Chem Int Ed.* 2013;52(45):11808–11812.

[98] Hoult DI, Richards RE. The signal-to-noise ratio of the nuclear magnetic resonance experiment. *J Magn Reson.* 1976;24(1):71–85.

[99] Bloch F. Nuclear induction. *Phys Rev.* 1946;70(7-8):460–474.

[100] McConnell HM. Reaction Rates by Nuclear Magnetic Resonance. *J Chem Phys.* 1958;28(3):430–431.

[101] Banci L, Luchinat C. Selective versus non-selective T1 experiments to determine metal-nucleus distances in paramagnetic metalloproteins. *Inorganica Chimica Acta.* 1998;276(1-2):373–379.

[102] Schanda P. Fast-pulsing longitudinal relaxation optimized techniques: Enriching the toolbox of fast biomolecular NMR spectroscopy. *Progress in Nuclear Magnetic Resonance Spectroscopy.* 2009;55(3):238–265.

[103] Favier A, Brutscher B. Recovering lost magnetization: polarization enhancement in biomolecular NMR. *J Biomol NMR.* 2011;49(1):9–15.

[104] Billeter M, Orekhov VY. *Novel Sampling Approaches in Higher Dimensional NMR.* Berlin: Springer, 1st ed. 2012.

[105] Hiller S, Fiorito F, Wüthrich K, Wider G. Automated projection spectroscopy (APSY). *Proc Natl Acad Sci USA.* 2005;102(31):10876–10881.

[106] Luan T, Jaravine V, Yee A, Arrowsmith CH, Orekhov VY. Optimization of resolution and sensitivity of 4D NOESY using multi-dimensional decomposition. *J Biomol NMR.* 2005;33(1):1–14.

[107] Tugarinov V, Kay LE, Ibraghimov I, Orekhov VY. High-resolution four-dimensional 1H-13C NOE spectroscopy using methyl-TROSY, sparse data acquisition, and multidimensional decomposition. *J Am Chem Soc.* 2005;127(8):2767–2775.

[108] Kazimierczuk K, Stanek J, Zawadzka-Kazimierczuk A, Koźmiński W. Random sampling in multidimensional NMR spectroscopy. *Progress in Nuclear Magnetic Resonance Spectroscopy.* 2010;57(4):420–434.

[109] Felli IC, Brutscher B. Recent Advances in Solution NMR: Fast Methods and Heteronuclear Direct Detection. *ChemPhysChem*. 2009;10(9-10):1356–1368.

[110] Szyperski T, Wider G, Bushweller JH, Wüthrich K. Reduced dimensionality in triple-resonance NMR experiments. *J Am Chem Soc*. 1993;115(20):9307–9308.

[111] Brutscher B, Morelle N, Cordier F, Marion D. Determination of an Initial Set of NOE-Derived Distance Constraints for the Structure Determination of15N/13C-Labeled Proteins. *J Magn Reson B*. 1995;109(2):238–242.

[112] Kupče E, Freeman R. Projection-reconstruction of three-dimensional NMR spectra. *J Am Chem Soc*. 2003;125(46):13958–13959.

[113] Kazimierczuk K, Orekhov VY. Accelerated NMR Spectroscopy by Using Compressed Sensing. *Angew Chem Int Ed*. 2011;50(24):5556–5559.

[114] Holland DJ, Bostock MJ, Gladden LF, Nietlispach D. Fast multidimensional NMR spectroscopy using compressed sensing. *Angew Chem Int Ed*. 2011;50(29):6548–6551.

[115] Kazimierczuk K, Zawadzka-Kazimierczuk A, Koźmiński W. Non-uniform frequency domain for optimal exploitation of non-uniform sampling. *Journal of Magnetic Resonance*. 2010;205(2):286–292.

[116] Neduva V, Russell RB. Linear motifs: Evolutionary interaction switches. *FEBS Letters*. 2005;579(15):3342–3345.

[117] Diella F, Haslam N, Chica C, Budd A, Michael S, Brown NP, Travé G, Gibson TJ. Understanding eukaryotic linear motifs and their role in cell signaling and regulation. *Front Biosci*. 2008;13:6580–6603.

[118] Felsani A, Mileo AM, Paggi MG. Retinoblastoma family proteins as key targets of the small DNA virus oncoproteins. *Oncogene*. 2006;25(38):5277–5285.

[119] Butel JS. Viral carcinogenesis: revelation of molecular mechanisms and etiology of human disease. *Carcinogenesis*. 2000;21(3):405–426.

[120] zur Hausen H. Papillomaviruses in anogenital cancer as a model to understand the role of viruses in human cancers. *Cancer Res*. 1989;49(17):4677–4681.

[121] zur Hausen H. Papillomavirus infections-a major cause of human cancers. *Biochim Biophys Acta*. 1996;1288(2):F55–78.

[122] Gillison ML, Koch WM, Capone RB, Spafford M, Westra WH, Wu L, Zahurak ML, Daniel RW, Viglione M, Symer DE, Shah KV, Sidransky D. Evidence for a causal association between human papillomavirus and a subset of head and neck cancers. *J Natl Cancer Inst*. 2000;92(9):709–720.

[123] Westra WH. The changing face of head and neck cancer in the 21st century: the impact of HPV on the epidemiology and pathology of oral cancer. *Head Neck Pathol*. 2009;3(1):78–81.

[124] Snow AN, Laudadio J. Human papillomavirus detection in head and neck squamous cell carcinomas. *Adv Anat Pathol*. 2010;17(6):394–403.

[125] Guan P, Howell-Jones R, Li N, Bruni L, de Sanjose S, Franceschi S, Clifford GM. Human papillomavirus types in 115,789 HPV-positive women: a meta-analysis from cervical infection to cancer. *Int J Cancer*. 2012;131(10):2349–2359.

[126] de Sanjose S, Quint WGV, Alemany L, Geraets DT, Klaustermeier JE, Lloveras B, Tous S, Felix A, Bravo LE, Shin HR, Vallejos CS, Alonso de Ruiz P, Lima MA, Guimera N, Clavero O, Alejo M, Llombart-Bosch A, Cheng-Yang C, Tatti SA, Kasamatsu E, Iljazovic E, Odida M, Prado R, Seoud M, Grce M, Usubutun A, Jain A, Suarez GAH, Lombardi LE, Banjo A, Menéndez C, Domingo EJ, Velasco J, Nessa A, Chichareon SCB, Qiao YL, Lerma E, Garland SM, Sasagawa T, Ferrera A, Hammouda D, Mariani L, Pelayo A, Steiner I, Oliva E, Meijer CJ, Al-Jassar WF, Cruz E, Wright TC, Puras A, Llave CL, Tzardi M, Agorastos T, Garcia-Barriola V, Clavel C, Ordi J, Andújar M, Castellsagué X, Sánchez GI, Nowakowski AM, Bornstein J, Muñoz N, Bosch FX. Human papillomavirus genotype attribution in invasive cervical cancer: a retrospective cross-sectional worldwide study. *Lancet Oncol*. 2010;11(11):1048–1056.

[127] Yang L, Mohr I, Fouts E, Lim DA, Nohaile M, Botchan M. The E1 protein of bovine papilloma virus 1 is an ATP-dependent DNA helicase. *Proc Natl Acad Sci USA*. 1993;90(11):5086–5090.

[128] Sverdrup F, Khan SA. Replication of human papillomavirus (HPV) DNAs supported by the HPV type 18 E1 and E2 proteins. *Journal of Virology*. 1993;68(1):505–509.

[129] Bouvard V, Storey A, Pim D, Banks L. Characterization of the human papillomavirus E2 protein: evidence of trans-activation and trans-repression in cervical keratinocytes. *EMBO J*. 1994;13(22):5451–5459.

[130] Wilson R, Fehrmann F, Laimins LA. Role of the E1–E4 protein in the differentiation-dependent life cycle of human papillomavirus type 31. *Journal of Virology*. 2005; 79(11):6732–6740.

[131] Fehrmann F, Klumpp DJ, Laimins LA. Human papillomavirus type 31 E5 protein supports cell cycle progression and activates late viral functions upon epithelial differentiation. *Journal of Virology*. 2003;77(5):2819–2831.

[132] Munger K, Phelps WC, Bubb V, Howley PM, Schlegel R. The E6 and E7 genes of the human papillomavirus type 16 together are necessary and sufficient for transformation of primary human keratinocytes. *Journal of Virology*. 1989;63(10):4417–4421.

[133] Band V, Zajchowski D, Swisshelm K, Trask D, Kulesa V, Cohen C, Connolly J, Sager R. Tumor progression in four mammary epithelial cell lines derived from the same patient. *Cancer Res*. 1990;50(22):7351–7357.

[134] Munger K, Phelps WC. The human papillomavirus E7 protein as a transforming and transactivating factor. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*. 1993;1155(1):111–123.

[135] Scheffner M, Huibregtse JM, Vierstra RD, Howley PM. The HPV-16 E6 and E6-AP complex functions as a ubiquitin-protein ligase in the ubiquitination of p53. *Cell*. 1993;75(3):495–505.

[136] Dyson N, Howley PM, Munger K, Harlow E. The human papilloma virus-16 E7 oncoprotein is able to bind to the retinoblastoma gene product. *Science*. 1989; 243(4893):934–937.

[137] Xue B, Ganti K, Rabionet A, Banks L, Uversky VN. Disordered interactome of human papillomavirus. *Curr Pharm Des*. 2013;20(8):1274–1292.

[138] Phelps WC, Munger K, Yee CL, Barnes JA, Howley PM. Structure-function analysis of the human papillomavirus type 16 E7 oncoprotein. *Journal of Virology*. 1992; 66(4):2418–2427.

[139] Barbosa MS, Lowy DR, Schiller JT. Papillomavirus polypeptides E6 and E7 are zinc-binding proteins. *Journal of Virology*. 1989;63(3):1404–1407.

[140] Uversky VN, Roman A, Oldfield CJ, Dunker AK. Protein Intrinsic Disorder and Human Papillomaviruses: Increased Amount of Disorder in E6 and E7 Oncoproteins from High Risk HPVs. *J Proteome Res*. 2006;5(8):1829–1842.

[141] Liu X, Clements A, Zhao K, Marmorstein R. Structure of the human Papillomavirus E7 oncoprotein and its mechanism for inactivation of the retinoblastoma tumor suppressor. *J Biol Chem.* 2006;281(1):578–586.

[142] Ohlenschläger O, Seiboth T, Zengerling H, Briese L, Marchanka A, Ramachandran R, Baum M, Korbas M, Meyer-Klaucke W, Dürst M, Görlach M. Solution structure of the partially folded high-risk human papilloma virus 45 oncoprotein E7. *Oncogene.* 2006;25(44):5953–5959.

[143] Lee JO, Russo AA, Pavletich NP. Structure of the retinoblastoma tumour-suppressor pocket domain bound to a peptide from HPV E7. *Nature.* 1998;391(6670):859–865.

[144] Chemes LB, Sanchez IE, Smal C, de Prat-Gay G. Targeting mechanism of the retinoblastoma tumor suppressor by a prototypical viral oncoprotein. *FEBS Journal.* 2010;277(4):973–988.

[145] Jewers RJ, Hildebrandt P, Ludlow JW, Kell B, McCance DJ. Regions of human papillomavirus type 16 E7 oncoprotein required for immortalization of human keratinocytes. *Journal of Virology.* 1992;66(3):1329–1335.

[146] Huh K, Zhou X, Hayakawa H, Cho JY, Libermann TA, Jin J, Harper JW, Munger K. Human Papillomavirus Type 16 E7 Oncoprotein Associates with the Cullin 2 Ubiquitin Ligase Complex, Which Contributes to Degradation of the Retinoblastoma Tumor Suppressor. *Journal of Virology.* 2007;81(18):9737–9747.

[147] Berezutskaya E, Bagchi S. The human papillomavirus E7 oncoprotein functionally interacts with the S4 subunit of the 26 S proteasome. *J Biol Chem.* 1997; 272(48):30135–30140.

[148] Davies R, Hicks R, Crook T, Morris J, Vousden K. Human papillomavirus type 16 E7 associates with a histone H1 kinase and with p107 through sequences necessary for transformation. *Journal of Virology.* 1993;67(5):2521–2528.

[149] Dyson N, Guida P, Munger K, Harlow E. Homologous sequences in adenovirus E1A and human papillomavirus E7 proteins mediate interaction with the same set of cellular proteins. *Journal of Virology.* 1992;66(12):6893–6902.

[150] Funk JO, Waga S, Harry JB, Espling E, Stillman B, Galloway DA. Inhibition of CDK activity and PCNA-dependent DNA replication by p21 is blocked by interaction with the HPV-16 E7 oncoprotein. *Genes & Development.* 1997;11(16):2090–2100.

[151] K ZT, W Z, Mannhardt B, R T, JW B, P JD. Inactivation of the cdk inhibitor p27KIP1 by the human papillomavirus type 16 E7 oncoprotein. *Oncogene*. 1996; 13(11):2323–2330.

[152] Roman A, Munger K. The papillomavirus E7 proteins. *Virology*. 2013;445(1-2):138–168.

[153] Ginsberg HS. The life and times of adenoviruses. *Adv Virus Res*. 1999;54:1–13.

[154] Gallimore PH. Tumour production in immunosuppressed rats with cells transformed in vitro by adenovirus type 2. *J Gen Virol*. 1972;16(1):99–102.

[155] Dimmock NJ, Easton AJ, Leppard KN. *Introduction to Modern Virology*. John Wiley & Sons, 6th ed. 2009.

[156] Chroboczek J, Bieber F, Jacrot B. The sequence of the genome of adenovirus type 5 and its comparison with the genome of adenovirus type 2. *Virology*. 1992;186(1):280–285.

[157] Halldén G. Adenoviral gene expression and replication in human tumor explant models. *Methods Mol Biol*. 2011;797:65–78.

[158] Acheson NH. *Fundamentals of Molecular Virology*. Wiley, 2nd ed. 2011.

[159] Nevins JR, Ginsberg HS, Blanchard JM, Wilson MC, Darnell JE. Regulation of the primary expression of the early adenovirus transcription units. *Journal of Virology*. 1979;32(3):727–733.

[160] Jones N, Shenk T. An adenovirus type 5 early gene function regulates expression of other early viral genes. *Proc Natl Acad Sci USA*. 1979;76(8):3665–3669.

[161] Miller MS, Pelka P, Fonseca GJ, Cohen MJ, Kelly JN, Barr SD, Grand RJA, Turnell AS, Whyte P, Mymryk JS. Characterization of the 55-Residue Protein Encoded by the 9S E1A mRNA of Species C Adenovirus. *Journal of Virology*. 2012;86(8):4222–4233.

[162] Kimelman D, Miller JS, Porter D, Roberts BE. E1a regions of the human adenoviruses and of the highly oncogenic simian adenovirus 7 are closely related. *Journal of Virology*. 1985;53(2):399–409.

[163] Romero P, Obradovic Z, Dunker AK. Natively disordered proteins: functions and predictions. *Appl Bioinformatics*. 2004;3(2-3):105–113.

[164] Bayley ST, Mymryk JS. Adenovirus e1a proteins and transformation (review). *Int J Oncol*. 1994;5(3):425–444.

[165] Pelka P, Fonseca GJ, Yousef AF, Mymryk JS. Intrinsic Structural Disorder in Adenovirus E1A: a Viral Molecular Hub Linking Multiple Diverse Processes. *Journal of Virology*. 2008;82(15):7252–7263.

[166] Hu X, Lazar MA. The CoRNR motif controls the recruitment of corepressors by nuclear hormone receptors. *Nature*. 1999;402(6757):93–96.

[167] Arany Z, Newsome D, Oldread E, Livingston DM, Eckner R. A Family of Transcriptional Adapter Proteins Targeted by the E1a Oncoprotein. *Nature*. 1995; 374(6517):81–84.

[168] Deleu L, Shellard S, Alevizopoulos K, Amati B, Land H. Recruitment of TRRAP required for oncogenic transformation by E1A. *Oncogene*. 2001;20(57):8270–8275.

[169] Fuchs M, Gerber J, Drapkin R, Sif S, Ikura T, Ogryzko V, Lane WS, Nakatani Y, Livingston DM. The p400 complex is an essential E1A transformation target. *Cell*. 2001;106(3):297–307.

[170] Reid JL, Bannister AJ, Zegerman P, Martinez-Balbas MA, Kouzarides T. E1A directly binds and regulates the P/CAF acetyltransferase. *EMBO J*. 1998;17(15):4469–4477.

[171] Lipinski KS, Esche H, Brockmann D. Amino acids 1-29 of the adenovirus serotypes 12 and 2 E1A proteins interact with rap30 (TFIIF) and TBP in vitro. *Virus Research*. 1998;54(1):99–106.

[172] Turnell AS, Grand RJA, Gorbea C, Zhang X, Wang WL, Mymryk JS, Gallimore PH. Regulation of the 26S proteasome by adenovirus E1A. *EMBO J*. 2000;19(17):4759–4773.

[173] O'Connor MJ, Zimmermann H, Nielsen S, Bernard HU, Kouzarides T. Characterization of an E1A-CBP interaction defines a novel transcriptional adapter motif (TRAM) in CBP/p300. *Journal of Virology*. 1999;73(5):3574–3581.

[174] Liu X, Marmorstein R. Structure of the retinoblastoma protein bound to adenovirus E1A reveals the molecular basis for viral oncoprotein inactivation of a tumor suppressor. *Genes & Development*. 2007;21(21):2711–2716.

[175] Ikeda MA, Nevins JR. Identification of distinct roles for separate E1A domains in disruption of E2F complexes. *Mol Cell Biol.* 1993;13(11):7029–7035.

[176] Whyte P, Buchkovich KJ, Horowitz JM, Friend SH, Raybuck M, Weinberg RA, Harlow E. Association between an oncogene and an anti-oncogene: the adenovirus E1A proteins bind to the retinoblastoma gene product. *Nature.* 1988;334(6178):124–129.

[177] Hateboer G, Hijmans EM, Nooij JB, Schlenker S, Jentsch S, Bernards R. mUBC9, a novel adenovirus E1A-interacting protein that complements a yeast cell cycle defect. *J Biol Chem.* 1996;271(42):25906–25911.

[178] Zhang X, Turnell AS, Gorbea C, Mymryk JS, Gallimore PH, Grand RJA. The targeting of the proteasomal regulatory subunit S2 by adenovirus E1A causes inhibition of proteasomal activity and increased p53 expression. *J Biol Chem.* 2004; 279(24):25122–25133.

[179] Hateboer G, Gennissen A, Ramos YF, Kerkhoven RM, Sonntag-Buck V, Stunnenberg HG, Bernards R. BS69, a novel adenovirus E1A-associated protein that inhibits E1A transactivation. *EMBO J.* 1995;14(13):3159–3169.

[180] Johnson ES. Protein modification by SUMO. *Annu Rev Biochem.* 2004;73:355–382.

[181] Lillie JW, Green M, Green MR. An adenovirus E1a protein region required for transformation and transcriptional repression. *Cell.* 1986;46(7):1043–1051.

[182] Kuppuswamy MN, Chinnadurai G. Relationship between the transforming and transcriptional regulatory functions of adenovirus 2 E1a oncogene. *Virology.* 1987; 159(1):31–38.

[183] Mymryk JS, Shire K, Bayley ST. Induction of apoptosis by adenovirus type 5 E1A in rat cells requires a proliferation block. *Oncogene.* 1994;9(4):1187–1193.

[184] Jones N, Shenk T. Isolation of adenovirus type 5 host range deletion mutants defective for transformation of rat embryo cells. *Cell.* 1979;17(3):683–689.

[185] Pelka P, Torchia J, Turnell AS, Grand RJA, Mymryk JS. Transcriptional control by adenovirus E1A conserved region 3 via p300/CBP. *Nucleic Acids Research.* 2009; 37(4):1095–1106.

[186] Ablack JNG, Cohen M, Thillainadesan G, Fonseca GJ, Pelka P, Torchia J, Mymryk JS. Cellular GCN5 is a novel regulator of human adenovirus E1A-conserved region 3 transactivation. *Journal of Virology*. 2012;86(15):8198–8209.

[187] Rasti M, Grand RJA, Yousef AF, Shuen M, Mymryk JS, Gallimore PH, Turnell AS. Roles for APIS and the 20S proteasome in adenovirus E1A-dependent transcription. *EMBO J*. 2006;25(12):2710–2722.

[188] Chinnadurai G. CtBP, an unconventional transcriptional corepressor in development and oncogenesis. *Molecular Cell*. 2002;9(2):213–224.

[189] Zhang Z, Smith MM, Mymryk JS. Interaction of the E1A oncoprotein with Yak1p, a novel regulator of yeast pseudohyphal differentiation, and related mammalian kinases. *Mol Biol Cell*. 2001;12(3):699–710.

[190] Köhler M, Görlich D, Hartmann E, Franke J. Adenoviral E1A protein nuclear import is preferentially mediated by importin alpha3 in vitro. *Virology*. 2001;289(2):186–191.

[191] Ferreon JC, Martinez-Yamout MA, Dyson HJ, Wright PE. Structural basis for subversion of cellular control mechanisms by the adenoviral E1A oncoprotein. *Proc Natl Acad Sci USA*. 2009;106(32):13260–13265.